Stochastics Summer Term 2019 Cheat Sheets

TUHH Class

27th August 2019 3:18pm GMT

Contents

1 Thesaurus / Definitions

1.1 Probabillity Theory

Random Experiment: A *random experiment* is an experimental procedure with a definite set of possible outcomes but an unpredictable single outcome.

sample space Ω: The set of possible outcomes of a random experiment. In general, any $\Omega \neq \emptyset$ is a *sample space*.

sample: A (random) *sample* of length n is a list of single outcomes of n realizations of the experiment: $(x_1, \ldots, x_n), with\, x_1, \ldots, x_n \in \Omega$. Outcomes can occur multiple times.

Power set: The set of all subsets of any set Ω, including the empty set and Ω itself, is called the *power set*, also denoted as $\mathcal{P}(\Omega)$.

σ-algebra (σ-field): Let Ω be a sample space. Then $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ is a σ*-algebra* on Ω if:

1. $\emptyset, \Omega \in \mathcal{A}$,
2. $A \in \mathcal{A} \implies \Omega \setminus A \in \mathcal{A}$,
3. $(A_n) \in \mathcal{A} \implies \bigcup A_n \in \mathcal{A}$.

For some $\mathcal{G} \subseteq \mathcal{P}(\Omega)$, the unique smallest σ-algebra $\mathcal{A}$ containing $\mathcal{G}$ is called the *σ-algebra generated by $\mathcal{G}$*. It is denoted by $\sigma(\mathcal{G}) := \mathcal{A}$.

In the discrete case, the σ-algebra is given by $\mathcal{A} := \mathcal{P}(\Omega)$.

Borel σ-algebra on $\mathbb{R}^n$: Let $\sigma = \mathbb{R}^n, \mathcal{G} = \{\prod_{j=1}^n (a_j, b_j]; a_j, b_j \in \mathbb{Q}, a_j < b_j \, (j \in \{1, \dots, n\})\}$. Then the Borel σ-algebra $\mathcal{B}(\mathbb{R}^n)$ on $\mathbb{R}^n$ is defined as $\mathcal{B}(\mathbb{R}^n) := \sigma(\mathcal{G})$. The Borel σ-algebra on subsets $\Omega \subseteq \mathbb{R}^n$ is defined as $\mathcal{B}(\Omega) := \{A \cap \Omega; A \in \mathcal{B}(\mathbb{R}^n)\}$.

In the absolutely continuous case, the sample space is a Borel set (so $\Omega \in \mathcal{B}(\mathbb{R}^n)$) and the σ-algebra is given by the Borel σ-algebra on Ω, i.e. $\mathcal{A} := \mathcal{B}(\Omega)$.

Event Space, Event: Let $\mathcal{A}$ be σ-algebra on Ω. Then $(\Omega, \mathcal{A})$ is an *event space.* Every $A \in \mathcal{A}$ is an *event.*

σ-algebra generated by $\mathcal{G}$: Let Ω be a sample space and $\mathcal{G} \subseteq \mathcal{P}(\Omega)$. Then the unique smallest σ-algebra $\mathcal{A}$ on Ω containing $\mathcal{G}$ is called *σ-algebra generated by* $\mathcal{G}$ and denoted by $\sigma(\mathcal{G}) := \mathcal{A}$

Probability Measure, Probability Space: Let Ω be a sample space, $\mathcal{A} \subseteq \mathcal{P}\,(\Omega)$ a σ-algebra on Ω. Then $\mathbb{P} : \mathcal{A} \to [0, 1]$ is called *probability measure* or *distribution* on Ω if:

1. $\mathbb{P}(\emptyset) = 0$,
2. $\mathbb{P}(\Omega) = 1$,
3. $(A_n)_{n \in \mathbb{N}} \in \mathcal{A}$ pairwise disjoint ($A_n \cap A_k = \emptyset$ for $n \neq k$) $\implies \mathbb{P}(\bigcup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \mathbb{P}(A_n)$.

If $\mathbb{P}$ is a probability measure on Ω, then $(\Omega, \mathcal{A}, \mathbb{P})$ is called a *probability space.* (See p. 11, Ch. 1 and 2)

Probability Mass Function: Let Ω be an at most countable sample space (i.e. we're working with a discrete model). For $\omega \in \Omega$ let $p_\omega \in \mathbb{R}$. Then $(p_\omega)_{\omega \in \Omega}$ is a probability mass function (pmf) if

1. $p_\omega \geq 0$ for all $\omega \in \Omega$
2. $\sum_{\omega \in \Omega} p_\omega = 1$.

Corresponding probability measures and probability mass functions can be obtained using the probabilities of singleton events ($p_\omega = \mathbb{P}(\{\omega\})$.

Random Variable and Random Vector: Let $(\Omega, \mathcal{A})$ be an event space.

(a) Let $X : \Omega \to \mathbb{R}$. Then X is called a *random variable* provided $X^{-1}(B) \in \mathcal{A}$ for all $B \in \mathcal{B}(\mathbb{R})$.

(b) Let $X : \Omega \to \mathbb{R}^n$. Then X is called a *random vector* or *random variable* provided $X^{-1}(B) \in \mathcal{A}$ for all $B \in \mathcal{B}(\mathbb{R}^n)$.

Let $(\Omega_1, \mathcal{A}_1, \mathbb{P}_1)$, $(\Omega_2, \mathcal{A}_2, \mathbb{P}_2)$ be probability spaces and X_1 and X_2 random variables on Ω_1 and Ω_2 respectively. Then X_1 and X_2 are called identically distributed provided $(\mathbb{P}_1)_{X_1} = (\mathbb{P}_2)_{X_2}$

Distribution Function for Random Variables: Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X a random variable. Then the probability measure $\mathbb{P}_X : \mathcal{B}(\mathbb{R}) \to [0, 1]$,

$$\mathbb{P}_X(A) := \mathbb{P}(X \in A) := \mathbb{P}(\{X \in A\}) = \mathbb{P}(X^{-1}(A)) \quad (\text{for } A \in \mathcal{B}(\mathbb{R}))$$

is called the distribution of X. If $\tilde{\mathbb{P}}$ is a distribution (i.e. a probability measure) on $\mathbb{R}$, we write $X \sim \tilde{\mathbb{P}} =$ if $\mathbb{P}_X = \tilde{\mathbb{P}}$. Analogously, we define distributions for random vectors.

Cumulative Distribution Function for Random Variables: Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X a random variable. Then $F_X : \mathbb{R} \to [0,1]$ defined by

$F_X(x) := \mathbb{P}_X((-\infty, x]) = \mathbb{P}(X \leq x)$ $(x \in \mathbb{R})$ is called (cumulative) distribution function of X.

Expectation of a Random Variable: Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, X an integrable random variable or random vector. Then we define the expectation of X by

$$\mathbb{E}(X) := \int X d\mathbb{P} = \int_\Omega id(X)\, d\mathbb{P} = \int_\mathbb{R} id\, d\mathbb{P}_X = \int_\mathbb{R} x\, d\mathbb{P}_X(x).$$

Conditional Expectation w.r.t. an Event: Let X be a random variable, $B \in \mathcal{A}$ with $\mathbb{P}(B) > 0$. Then

$$\mathbb{E}(X|B) := \frac{\mathbb{E}(\mathbb{1}_B X)}{\mathbb{P}(B)} = \int X\, d\mathbb{P}(\cdot|B)$$

is called conditional expectation of X with respect to B.

Variance of a Random Variable: Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, X a random variable such that X^2 integrable. Then

$$\mathrm{Var}(X) := \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2.$$

Let X,Y be independent random variables, X^2, Y^2 integrable, a $\in \mathbb{R}$. Then

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) \quad \text{and} \quad \mathrm{Var}(\alpha X) = \alpha^2 \,\mathrm{Var}(X).$$

Covariance and Correlation Coefficient of two Random Variables: Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, X, Y random variables such that X^2, Y^2 are integrable. Then

$$\mathrm{Cov}(X, Y) := \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

is called *covariance* of X and Y. Moreover,

$$\varrho(X, Y) := \frac{\mathrm{Cov}(X, Y)}{\sigma(X)\sigma(Y)}$$

is called *correlation coefficient* of X and Y. X and Y are *uncorrelated* provided $\varrho(X, Y) = 0$.

(3.44 Lemma:) Let X,Y be normally distributed random variables, X,Y uncorrelated. Then X and Y are independent.

Cov(X,Y) = $\mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$. Covariance and Independence. If two random variables are independent, then they are uncorrelated. The inverse is not necessarily true.

If X and Y are independent $\to Cov(X, Y) = 0 \to \mathbb{E}(XY) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$.

The variance of a sum of random variables can be found by:

$$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$$

$$Var(\sum_{i=1}^{n}(X_i)) = \sum_{i=1}^{n} Var(X_i) + 2\sum_{i<j}^{n} Cov(X_i, X_j)$$

If X and Y are independent then they have covariance 0, so Var(X+Y) = Var(X) + Var(Y).

General Moments: Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, X a random variable, $k \in \mathbb{N}$. If X is integrable, we call

$$m_k(X) := \mathbb{E}(X^k)$$

the kth moment of X.

Moment Generating Function: Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, X a random variable. For $t \in \mathbb{R}$ we define

$$M_X(t) := \mathbb{E}(e^{tX}),$$

provided the expectation exists. The function M_X is called moment generating function of X.

Characteristic Function: Let X be a random variable. For $t \in \mathbb{R}$ we define

$$\varphi_X(t) := \mathbb{E}(e^{itX}),$$

provided the expectation exists. The function $\varphi_X(t)$ is called characteristic function of X.

Let X,Y be random variables. Let X, Y be independent a,b $\in \mathbb{R}$. Then

$$\varphi_{aX+bY}(t) = \varphi_X(at) \cdot \varphi_Y(bt),$$

for all t $\in \mathbb{R}$ where all functions are defined and if

$$\varphi_X = \varphi_Y,$$

then

$$\mathbb{P}_X = \mathbb{P}_Y,$$

i.e. X and Y have the same distribution (and therefore also the same cumulative distribution function).

Conditional Probability (Prop 2.32): Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $B \in \mathcal{A}$ with $\mathbb{B} > 0$. Then $\mathbb{P}(\cdot|B) : \mathcal{A} \to [0,1]$, $\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ defines the unique probability measure on Ω satisfying $\mathbb{P}(B|B) = 1$ and for sub-events of B the measure $\mathbb{P}(\cdot|B)$ is proportional to $\mathbb{P}$, i.e. there exists $c_B > 0$ such that for all $A \in \mathcal{A}$ with $A \subseteq B$ we have $\mathbb{P}(A|B) = c_B \mathbb{P}(A)$.

Independence of a Family of Events: Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, I a non-empty index set and $(A_\iota \in I)$ in $\mathcal{A}$. Then $(A_\iota)_{\iota \in I}$ is called *independent* if for all non-empty finite subsets $\emptyset \neq J \subseteq I$ we have

$$\mathbb{P}\Big(\bigcap_{\iota \in J} A_\iota\Big) = \prod_{\iota \in J} \mathbb{P}(A_\iota).$$

Independent Random Variables: Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, X, Y random variables. Then X and Y are called independent if and only if

$$F_{(X,Y)}(x, y) = F_X(x) \cdot F_Y(y) \quad (x, y \in \mathbb{R}).$$

1.2 Statistics

Sample Mean/Empirical Average: Let $\Omega = \mathbb{R}^d$ and $(x_1, \ldots, x_n)$ a sample. Then the *sample mean* $\bar{x}$ is defined as:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Properties:

1. linear: $\overline{\alpha x + \beta y} = \alpha \overline{x} + \beta \overline{y}$.

Median: If we assume that $(x_1, \ldots, x_n)$ is increasing, i.e. $x_k \leq x_{k+1}$ for $k \in \{1, \ldots, n-1\}$, then

$$m_x := \begin{cases} x_{(n+1)/2} & \text{n is odd,} \\ \frac{1}{2}(x_{n/2} + x_{n/2+1}) & \text{n is even} \end{cases}$$

is called *median*.

Sample Correlation: Let $x := (x_1, \ldots, x_n)$ and $y := (y_1, \ldots, y_n)$ be two samples. Then

$$r := \frac{\sum_{k=1}^{n} (x_k - \bar{x})(y_k - \bar{x})}{\sqrt{\sum_{k=1}^{n} (x_k - \bar{x})^2} \sqrt{\sum_{k=1}^{n} (y_k - \bar{y})^2}} = \frac{\sum_{k=1}^{n} x_k y_k - n\bar{x}\bar{y}}{(n-1) s_x s_y}$$

Empirical Variance: Let $\Omega = \mathbb{R}^d$ and $(x_1, \ldots, x_n)$ a sample. Then the sample's *empirical variance* $\bar{x}$ is defined as:

$$\text{Var}(x) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2$$

Properties:

1. $\text{Var}(\alpha x + b) = \alpha^2 \cdot \text{Var}(x)$
2. $\text{Var}(x) = \overline{x^2} - (\overline{x})^2$.

Sample Space (Statistics): Let $\mathbb{X}_n \subseteq \mathbb{R}^n$ be the set of all possible values a sample of X of size n can have. Then $\mathbb{X}_n$ is called sample space.

Set of possible Distributions, Parameter Space: Let $\{\mathbb{P}_\vartheta; \vartheta \in \Theta\}$ be a set of probability measures on $\mathbb{R}$, i.e. the set of possible distributions X can have, where ϑ is some parameter. Then $\{\mathbb{P}_\vartheta; \vartheta \in \Theta\}$ is called set of possible distributions. The set Θ is called *parameter space*.

Statistical Model: Let $\mathbb{X}_n$ be a sample space, $\{\mathbb{P}_\vartheta; \vartheta \in \Theta\}$ a set of possible distributions. Let $\mathbb{P}_\vartheta^n$ be the n-fold product measure of $\mathbb{P}_\vartheta$. Then $(\mathbb{X}_n, \mathcal{B}(\mathbb{X}_n), \{\mathbb{P}_\vartheta^n; \vartheta \in \Theta\})$ is called a statistical model.

1.2.1 Point Estimators

Let $(\mathbb{X}_n, \mathcal{B}(\mathbb{X}_n), \{\mathbb{P}_\vartheta^n; \vartheta \in \Theta\})$ be a statistical model.

Point Estimator: Let $(\mathbb{X}_n, \mathcal{B}(\mathbb{X}_n), \{\mathbb{P}_\vartheta^n; \vartheta \in \Theta\})$ be a statistical model. Let $T_n : \mathbb{X}_n \to \Theta$ be a random variable. Then T_n is called *point estimator*.

Bias of a Point Estimator: Let T_n be a point estimator for $\vartheta \in \Theta$. Then $bias_\vartheta(T_n) := \mathbb{E}_\vartheta^n(T_n) - \vartheta$ is called *bias* of T_n. T_n is called *unbiased* if $bias_\vartheta(T_n) = 0$ for all $\vartheta \in \Theta$.

Lemma 6.6: (a) Let X be integrable. Then the sample mean is an unbiased point estimator for the mean.

(b) Let X^2 be integrable. Then the empirical variance is an unbiased point estimator for the variance.

Consistency of a Sequence of Point Estimators: For $n \in \mathbb{N}$ let T_n be a point estimator on $\mathbb{X}_n$ for $\vartheta \in \Theta$. Then $(T_n)_{n \in \mathbb{N}}$ is said to be *consistent* provided for all $\vartheta \in \Theta$ and $\epsilon \in (0,1)$ we have

$$\lim_{n \to \infty} \mathbb{P}_\vartheta^{\mathbb{N}}(|T_n - \vartheta| > \epsilon) = 0.$$

Lemma 6.8: (a) Let X be integrable. Then the sample mean is consistent.

(b) Let X^2 be integrable. Then the empirical variance is consistent.

Domination of Estimators, Relative Efficiency: Let T_n, $\tilde{T}_n$ be two unbiased estimators for $\vartheta \in \Theta$. Then T_n is said to *dominate* $\tilde{T}_n$ if

$$\mathrm{Var}^n_\vartheta(T_n) = \mathbb{E}^n_\vartheta((T_n - \vartheta)^2) \leq \mathbb{E}^n_\vartheta((\tilde{T}_n - \vartheta)^2) = \mathrm{Var}^n_\vartheta(\tilde{T}_n)$$

for all $\vartheta \in \Theta$, and there exists one $\vartheta \in \Theta$ where the inequality is strict. The *relative efficiency* of T_n and $\tilde{T}_n$ is defined by

$$e(T_n, \tilde{T}_n)(\vartheta) := \frac{\mathbb{E}^n_\vartheta((\tilde{T}_n - \vartheta)^2)}{\mathbb{E}^n_\vartheta(T_n - \vartheta)^2)} = \frac{\mathrm{Var}^n_\vartheta(\tilde{T}_n)}{\mathrm{Var}^n_\vartheta(T_n)}.$$

Lemma 6.9: Let $X \sim \texttt{Normal}(\mu, \sigma^2)$. Then the most efficient unbiased estimator for the mean $\mathbb{E}(X)$ is given by the sample mean.

Likelihood Function and Maximum Likelihood Estimator: Let $(\mathbb{X}_n, \mathcal{B}(\mathbb{X}_n), \{\mathbb{P}^n_\vartheta; \vartheta \in \Theta\})$ be a statistical model. For all $\vartheta \in \Theta$ let $\mathbb{P}^n_\vartheta$ be either discrete with pmf $(f_\vartheta(x))_x$ or (absolutely) continuous with pdf f_ϑ. Then $L : \mathbb{X}_n \times \Theta \to [0, \infty), L(x, \vartheta) := f_\vartheta(x)$ is called *likelihood function* for the model. A point estimator $T_n : \mathbb{X}_n \to \Theta$ is called *maximum likelihood estimator* if

$$L(x, T_n(x)) = \max_{\vartheta \in \Theta} L(x, \vartheta) \quad (x \in \mathbb{X}_n).$$

1.2.2 Interval Estimators

Let $(\mathbb{X}_n, \mathcal{B}(\mathbb{X}_n), \{\mathbb{P}^n_\vartheta; \vartheta \in \Theta\})$ be a statistical model.

Interval Estimator, Confidence Interval, Error Level, Confidence Level: Let $\alpha \in (0, 1)$. Then $C : \mathbb{X}_n \to \mathcal{P}(\mathbb{R})$ is called *interval estimator* if $C(x)$ is closed interval for all $x \in \mathbb{X}_n$ and:

$$\mathbb{P}^n_\vartheta(\{x \in \mathbb{X}_n; \vartheta \in C(x)\}) \geq 1 - \alpha \quad \text{for all } \vartheta \in \Theta.$$

Then $C(x)$ is *confidence interval* for the sample x. Here, α is called *error level* and $1 - \alpha$ is called *confidence level*.

Quantile: Let Q be a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R})), \alpha \in (0, 1)$. Then $z \in \mathbb{R}$ is called $\alpha -$ *quantile* of Q if

$$Q((-\infty, z]) \geq \alpha \text{ and } Q([z, \infty)) \geq 1 - \alpha.$$

1.2.3 Statistical Hypothesis Testing

Statistical Hypothesis Testing: 1. Formulation of the statistical model $(\mathbb{X}_n, \mathcal{B}(\mathbb{X}_n), \{\mathbb{P}^n_\vartheta; \vartheta \in \Theta\})$.

2. Formulation of null hypothesis and alternative: Decompose Θ into Θ_0 and Θ_1. Then formulate the *null hypothesis* $H_0 : \vartheta \in \Theta_0$ and the *alternative* $H_1 : \vartheta \in \Theta_1$.

3. Choice of the significance level: There are two possible sources of error:

   — type I error: H_0 is true, but will be rejected by means of the sample (a false rejection of H_0)

   — type II error: H_1 is false, but will not be rejected by means of the sample (a false acceptance of H_0)

   We choose a *significance level* $\alpha \in (0, 1)$ such that the probability of getting the type I error is at most α.

4. Choice of a decision rule: For a *test*, i.e. a random variable $T_n : \mathbb{X}_n \to \mathbb{R}$, find the *critical region* $K \subseteq \mathbb{R}$ such that
$$\max_{\vartheta \in \Theta_0} \mathbb{P}_\vartheta^n(T_n \in K) \leq \alpha.$$
5. Perform the test: Let now $x = (x_1, \ldots, x_n) \in \mathbb{X}_n$ be a sample. Calculate $T_n(x) \in \mathbb{R}$.
   — If $T_n(x) \notin K$ then accept H_0.
   — If $T_n(x) \in K$ then reject H_0.

2 Formulas / Theorems

2.1 Stochastic

Lemma 2.2: Let $(\Omega, \mathcal{A})$ be an event space, $A, B \in \mathcal{A}$. Then $A \cup B, A \cap B, A \setminus B \in \mathcal{A}$. Moreover, if $(A_n)_{n \in \mathbb{N}}$ in $\mathcal{A}$ then also $\bigcap_{n \in \mathbb{N}} A_n \in \mathcal{A}$.

Bayes: Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $A, B \in \mathcal{A}$ with $\mathbb{P}(A), \mathbb{P}(B) > 0$. Then
$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(B)\mathbb{P}(A \mid B)}{\mathbb{P}(A)}.$$

Total probability: Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $n \in \mathbb{N}$, $B_1, \ldots, B_n \in \mathcal{A}$ pairwise disjoint such that $\Omega = \bigcup_{k=1}^n B_k$ and $\mathbb{P}(B_k) > 0$ for all $k \in \{1, \ldots, n\}$. Then
$$\mathbb{P}(A) = \sum_{k=1}^n \mathbb{P}(B_k)\mathbb{P}(A \mid B_k) \quad (A \in \mathcal{A}).$$

Operations on probability measures: Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space.

a) Let (A_n) in $\mathcal{A}$, $A_n \subseteq A_{n+1}$ for all $n \in \mathbb{N}$. Then $\mathbb{P}\left(\bigcup_{n=1}^\infty A_n\right) = \lim_{n \to \infty} \mathbb{P}(A_n)$.

b) Let (A_n) in $\mathcal{A}$, $A_{n+1} \subseteq A_n$ for all $n \in \mathbb{N}$. Then $\mathbb{P}\left(\bigcap_{n=1}^\infty A_n\right) = \lim_{n \to \infty} \mathbb{P}(A_n)$.

c) Let $A, B \in \mathcal{A}$, $A \subseteq B$. Then $\mathbb{P}(A) \leq \mathbb{P}(B)$.

Proposition 2.14: Let Ω be an at most countable sample space, $\mathcal{A} := \mathcal{P}(\Omega)$.

(a) Let $(p_\omega)_{\omega \in \Omega}$ be a pmf. Then $\mathbb{P} : \mathcal{A} \to [0,1], \mathbb{P}(A) := \sum_{\omega \in A} p_\omega$ defines a probability measure on Ω.

(b) Let $\mathbb{P} : \mathcal{A} \to [0,1]$ be a probability measure on Ω. Then $(p_\omega)_{\omega \in \Omega}$ defined by $p_\omega := \mathbb{P}(\{\omega\})$ is a pmf.

Multiplication rule: Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $n \in \mathbb{N}$ and $A_1, \ldots, A_n \in \mathcal{A}$. Then
$$\mathbb{P}(A_1 \cap \ldots \cap A_n) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2 \mid A_1) \cdot \ldots \cdot \mathbb{P}(A_n \mid A_1 \cap \ldots \cap A_{n-1}).$$

Convolution (Theorem 3.21): Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, X, Y be independent random variables, $a \in \mathbb{R} \setminus \{0\}$.

(a) Let X, Y be discrete with pmf p_X and p_Y, respectively. Then aX has the pmf p_{aX} given by $(p_{aX})_k := (p_X)_{\frac{k}{a}}$, and X, Y has the pmf p_{X+Y} given by the convolution $p_X * p_Y$, i.e.
$$(p_{X+Y})_k := \sum_l (p_X)_{k-l}(p_Y)_l.$$

(b) Let X, Y be (absolutely) continuous with pdf f_X and f_Y, respectively. Then aX has the pdf f_{aX} given by $(f_{aX})_k := \frac{1}{a} f_X(\frac{k}{a})$, and X, Y has the pdf f_{X+Y} given by the convolution $f_X * f_Y$, i.e.

$$f_{X+Y}(x) = \int_{-\infty}^{\infty} f_X(x-y) f_Y(y) \mathrm{d}y.$$

Lemma 3.36 (Linearity of Expectation): Let X, Y be integrable random variables, $a, b \in \mathbb{R}$. Then $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$. If X and Y are independent, then $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$

Lemma 3.39 (Variance of Independent Random Variables): Let X, Y be independent random variables, X^2, Y^2 integrable, $a \in \mathbb{R}$. Then $\operatorname{Var}(X+Y) = \operatorname{Var}(X) + \operatorname{Var}(Y)$ and $\operatorname{Var}(aX) = a^2 \operatorname{Var}(X)$.

Lemma 3.44: Let X, Y be normally distributed random variables, X, Y uncorrelated. Then X and Y are independent.

Lemma 3.45 (Moments/Expectation): Let X be a random variable, $k \in \mathbb{N}$. Then

$$m_k(X) = \mathbb{E}(X^k) = M_X^{(k)}(0)$$

provided the kth moment of X exists.

Proposition 3.48 (Moments): Let X, Y be random variables.

(a) Let X, Y be independent, $a, b \in \mathbb{R}$. Then

$$M_{aX+bY}(t) := M_X(at) M_Y(bt)$$

for all $t \in \mathbb{R}$ where all functions are defined.

(b) Let $M_X = M_Y$. Then $\mathbb{P}_X = \mathbb{P}_Y$, i.e. X and Y have the same distribution (and therefore the same cumulative distribution function).

Summing Independent RVs by Multiplying MGFs: If X and Y are independent, then

$$M_{X+Y}(t) := M_X(t) M_Y(t) = \mathbb{E}(e^{t(X+Y)}) = \mathbb{E}(e^{tX}) \cdot \mathbb{E}(e^{tY})$$

The MGF of the sum of two r.v.s is the product of the MGFs of those two r.v.s.

Lemma 3.7 (Random Variable): Let $(\Omega, \mathcal{A})$ be an event space, $X : \Omega \to \mathcal{R}$. Then

$$X \text{ is a random variable} \Leftrightarrow X^{-1}((a, b]) \in \mathcal{A} \text{ for all } a, b \in \mathbb{R}.$$

Theorem 3.10: Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X a random variable. Then $\mathbb{P}_X : \mathcal{B}(\mathbb{R}) \to [0, 1]$,

$$\mathbb{P}_X(A) := \mathbb{P}(X \in A) := \mathbb{P}(\{X \in A\}) = \mathbb{P}(X^{-1}(A)) \quad (A \in \mathcal{B}(\mathbb{R}))$$

defines a probabilility measure on $\mathbb{R}$.

Poisson Process: Let $(P_t)_{t \geq 0}$ be a continuous time discrete process taking values in $\mathbb{N}_0, \mu > 0$. Then $(P_t)_{t \geq 0}$ is called Poisson process with intensity μ if:

1. $\mathbb{P}(P_0 = 0) = 1$.
2. $(P_t)_{t \geq 0}$ has independent and stationary increments.
3. $P_t - P_s \sim \texttt{Poi}(\mu(t-s))$ for all s, t ≥ 0, s $<$ t.

The Poisson process is used to model number of occurrences of single events, like the number of incoming e-mails up to time t, or the number or produced failures up to time to.

We have a **Poisson process** of rate λ arrivals per unit time if the following conditions hold:

1. The number of arrivals in a time interval of length t is $\texttt{Poi}(\lambda t)$.
2. Numbers of arrivals in disjoint time intervals are independent.

If $0 \leq s \leq 1$ and $0 \leq k \leq n$ then $\mathbb{P}(P_s = k | P_1 = n) = \binom{n}{k} s^k (1-k)^{n-k}$.

$\mathbb{P}(P_s = k | P_1 = n) = \frac{\mathbb{P}(P_s = k \cap P_1 = n)}{\mathbb{P}(P_1 = n)}$

For a Poisson process with rate λ, the number of events in the interval (a,b] is Poisson with probability mass function:

$f(x) = \frac{(\lambda \cdot (b-a))^x e^{-\lambda(b-a)}}{x!}$ $(\lambda \geq 0)$.

$\mathbb{E}(P_{\lambda,t}) = \text{Var}\ (P_{\lambda,t}) = \lambda \cdot t$

Count-Time Duality Consider a Poisson process of emails arriving in an inbox at rate λ emails per hour. Let T_n be the time of arrival of the nth email (relative to some starting time 0) and N_t be the number of emails that arrive in $[0, t]$. Let's find the distribution of T_1. The event $T_1 > t$, the event that you have to wait more than t hours to get the first email, is the same as the event $N_t = 0$. So $T_1 \sim \texttt{Exp}(\lambda)$. By the memoryless property and similar reasoning, the interarrival times between emails are i.i.d. $\texttt{Exp}(\lambda)$, i.e., the differences $T_n - T_{n-1}$ are i.i.d. $\texttt{Exp}(\lambda)$.

Markov Process: Let $T \subseteq \mathbb{R}$ and $(X_t)_{t \in T}$ a discrete stochastic process. Then (X_t) is called *Markov process* if for all $n \in \mathbb{N}$, all $t_1, \ldots, t_n, t \in T$ such that $t_1 < t_2 < \ldots < t_n < t$, and all $x_1, x_2, \ldots, x_n, x \in \mathbb{R}$ we have

$$\mathbb{P}(X_t = x \mid X_{t_1} = x_1, \ldots, X_{t_n} = x_n) = \mathbb{P}(X_t = x \mid X_{t_n} = x_n).$$

Transition Probabilities: Let $(X_n)_{n \in \mathbb{N}_0}$ be a discrete Markov chain. Then

$$p_{x,y}(n, n+1) := \mathbb{P}(X_{n+1} = y \mid X_n = x)$$

is called *transition probability* from x at time n to y at time $n+1$. Analogously, we define

$$p_{x,y}(n, n+k) := \mathbb{P}(X_{n+k} = y \mid X_n = x)$$

for all $n, k \in \mathbb{N}_0$.

Lemma 4.6 (Markov Chain): Let $(X_n)_{n \in \mathbb{N}_0}$ be a discrete Markov chain. Then for the transition probabilities we have:

(a) $0 \leq p_{x,y}(n, n+1) \leq 1$ for all $n \in \mathbb{N}_0$ and $x, y \in \mathbb{R}$.

(b) If $\mathbb{P}(X_n = x) \neq 0$ then $\sum_y p_{x,y}(n, n+1) = 1$.

(c) Chapman-Kolmogorov equations: $p_{x,y}(n, n+k) = \sum_z p_{x,z}(n, n+l) p_{z,y}(n+l, n+k)$ for all $n, k, l \in \mathbb{N}_0, l \leq k, x, y \in \mathbb{R}$.

Time Homogeneity: Let $(X_n)_{n \in \mathbb{N}_0}$ be a discrete Markov chain. Then (X_n) is *time homogeneous* if $p_{x,n}(n, n+1)$ does not depend on n, i.e.

$$p_{x,y}(n, n+1) = \mathbb{P}(X_{n+1} = y \mid X_n = x) = \mathbb{P}(X_1 = y \mid X_0 = x) = p_{x,y}(0, 1)$$

for all $n \in \mathbb{N}_0$.

Transition Matrix: $(X_n)_{n \in \mathbb{N}_0}$ be a time-homogeneous discrete Markov chain. Let $p_{x,y} := p_{x,y}(0, 1)$. Then $P := (p_{x,y})_{x,y}$ is called transition matrix

Lemma 4.7: Let $(X_n)_{n \in \mathbb{N}_0}$ be a time-homogeneous discrete Markov chain with transition matrix $P, (p_x)_x$ the pmf of $X_0, n \in \mathbb{N}$. The pmf of X_n is given by $(p_x)_x P^n$.

2.1.1 Convergence of Random Variables

Markov's Inequality: Let X be a random variable, $p > 0$, $\epsilon > 0$. Then

$$\mathbb{P}(|X| \geq \epsilon) \leq \frac{\mathbb{E}(|X|^p)}{\epsilon^p}.$$

Chebyshev's Inequality: Let X be a random variable such that $|X|$ is integrable, $\epsilon > 0$. Then

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \epsilon) \leq \frac{\mathrm{Var}(X)}{\epsilon^2}.$$

Weak Law of Large Numbers: Let $(X_n)_{n\in\mathbb{N}}$ be independent and identically distributed (i.i.d.) random variables, $|X_1|$ integrable, $\mu := \mathbb{E}(X_1)$, $\epsilon > 0$. Then $(\frac{1}{n}\sum_{k=1}^n X_k)_n$ converges to a probabilty, i.e. for all $\epsilon > 0$ we have

$$\lim_{n\to\infty} \mathbb{P}\Big(\Big|\frac{1}{n}\sum_{k=1}^n X_k - \mu\Big| \geq \epsilon\Big) = 0.$$

Strong Law of Large Numbers: Let $(X_n)_{n\in\mathbb{N}}$ be independent and identically distributed (i.i.d.) random variables, $|X_1|^2$ integrable, $\mu := \mathbb{E}(X_1)$. Then $(\frac{1}{n}\sum_{k=1}^n X_k)_n$ converges to μ $\mathbb{P}$-almost-surely, i.e.

$$\mathbb{P}\Big(\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n X_k = \mu\Big) = 1.$$

Central Limit Theorem: Let $(X_n)_n \in \mathbb{N}$ be independent and identically distributed (i.i.d.) random variables, $|X_1|^2$ integrable, $\mu. = \mathbb{E}(X_1)$, $\sigma^2 := \mathrm{Var}(X_1)$. Then $\Big(\frac{\frac{1}{n}\sum_{k=1}^n X_k - \mu}{\frac{\sigma}{\sqrt{n}}}\Big)_n$ converges to the distribution $\texttt{Normal}(0, 1^2)$, i.e.

$$\lim_{n\to\infty} \mathbb{P}\Big(\frac{\frac{1}{n}\sum_{k=1}^n X_k - \mu}{\frac{\sigma}{\sqrt{n}}} \leq x\Big) = \Phi(x) \quad (x \in \mathbb{R}),$$

where $\Phi(x)$ is the cumulative distribution function of $\texttt{Normal}(0, 1^2)$.

2.2 Math

Series and Sums

Sum 1: $\sum\limits_{k=0}^{n-1} r^k = \frac{1-r^n}{1-r}$

Sum 2: $\sum\limits_{k=0}^{\infty} r^k = \frac{1}{1-r}$ $[\frac{d}{dr} : \sum\limits_{k=1}^{\infty} k \cdot r^{k-1} = \frac{1}{(1-r)^2}$ $]$

Sum 3: $\sum\limits_{k=0}^{n} \binom{n}{k} \cdot r^k = (1+r)^n$

Sum 4: $\sum\limits_{k=0}^{n} \binom{n}{k} \cdot a^{n-k} \cdot b^k = \sum\limits_{k=0}^{n} \binom{n}{k} (\frac{b}{a})^k \cdot a^n = (a+b)^n$

Sum 5: $\sum\limits_{k=r}^{\infty} \binom{k-1}{r-1} \cdot w^{k-r} = (1-w)^r$

Sum 6: $\sum\limits_{k=0}^{\infty} \binom{k+r-1}{r-1} \cdot w^k = \sum\limits_{k=0}^{\infty} \binom{k+r-1}{k} \cdot w^k = (1-w)^{-r}$ [Important for N.Bin(r,p)]

Sum 7: $\sum_{j=0}^{m} \binom{n+j}{n} = \binom{n+m+1}{n+1} = \binom{n+m+1}{m}$

Sum 8: $\sum_{k} \binom{r}{k}\binom{s}{n-k} = \binom{r+s}{n}$

Sum 9: $\sum_{i=0}^{m} (\binom{n}{i})^2 = \binom{2n}{n}$

Sum 10: $\sum_{k=0}^{\infty} \frac{x^k}{k!} = \lim_{k\to\infty}(1+\frac{x}{k})^k = e^x$

Derivative of Sum 10: $\frac{d}{dx}e^x = e^x = \sum_{k=0}^{\infty} k \cdot x^{k-1} \cdot \frac{1}{k!}$

Gauss Sum: $\sum_{j=1}^{n} j = \frac{n\cdot(n+1)}{2}$

Sum of Squares: $\sum_{j=1}^{n} j^2 = \frac{n\cdot(n+1)\cdot(2n+1)}{6}$

Finite Geometric Sum: $\sum_{j=m}^{n} x^j = \frac{x^m - x^{n+1}}{1-x}$, for $x \neq 1$

Infinite Geometric Sum: $\sum_{j=0}^{\infty} ax^j = \frac{a}{1-x}$, for $|x| < 1$.

Exponential Series: $\sum_{j=0}^{\infty} \frac{x^j}{j!} = e^x$

Binomial Coefficient: $\binom{n}{k} = \frac{n!}{k!(n-k)!} = \binom{n}{n-k}$ and $\binom{n}{n-1} = n$

$$k \cdot \binom{n}{k} = n \cdot \binom{n-1}{k-1}$$

$$\frac{n!}{n_1!n_2!(n-n_1-n_2)!} = \binom{n}{n_1} \cdot \binom{n-n_1}{n_2}$$

$$\binom{n}{2} = \frac{n\cdot(n-1)}{2}$$

use nCr for calculators

3 Distributions

3.1 Discrete Uniform Distribution

Prerequisites: Ω finite sample space, $\mathcal{A} := \mathcal{P}(\Omega)$
Name: $\mathbb{P} = \texttt{Uniform}(\Omega)$
Expectation: $\mathbb{E}(X) = \frac{a+b}{2}$ (for $\Omega = \{a, \ldots, b\}$)
Variance: $\texttt{Var}(X) = \frac{(b-a+1)^2-1}{12}$
MGF: $M_X(t) = \frac{e^{at}-e^{(b+1)t}}{n(1-e^t)}$
pmf: $p_\omega = \frac{1}{|\Omega|}$

explanation: Every element in the sample space Ω has the same probability to be drawn:

$\mathbb{P}(\omega) = \frac{1}{|\Omega|}, (\omega \in \Omega)$ (prerequisite; called Laplace-Experiment), $|\Omega|$ being the sample space size.

3.2 Bernoulli Distribution

Prerequisites: $n \in \mathbb{N}, \Omega = \{0,1\}^n, p \in [0,1]$
Name: $\mathbb{P} = \texttt{Bernoulli}(n,p)$
Expectation: $\mathbb{E}(X) = p$
Variance: $\texttt{Var}(X) = pq$
MGF: $M_X(t) = 1 - p + p * e^t$
pmf: $p_{(\omega_1,\dots,\omega_n)} := p^{\sum_{k=1}^n \omega_k}(1-p)^{\sum_{k=1}^n (1-\omega_k)}$

explanation: discrete. with replacement. Only two types (colors) exist. p is the probability of "success", getting the first type. Then, the probability of getting the other type must be 1-p. pmf describes the probability of getting exactly the sequence $\omega_1, \dots, \omega_n$.

3.3 Multinomial Distribution

Prerequisites: $n \in \mathbb{N}$, C finite set, $(p_c)_{c\in C}$ in $[0,1]$ with $\sum_{c\in C} p_c = 1$, $\Omega := \{(k_c)_{c\in C} \in \mathbb{N}_0^C; \sum_{c\in C} k_c = n\}$
Name: $\mathbb{P} = \texttt{Multin}(n,(p_c)_{c\in C})$
Expectation: $\mathbb{E}(X_i) = n * p_i$
Variance: $\texttt{Var}(X_i) = np_i(1-p_i)$
MGF: $M_X(t) = (\sum_{i=1}^k p_i * e^{t_i})^n$
pmf: $p_{(k_c)_{c\in C}} := \binom{n}{(k_c)_{c\in C}} \prod_{c\in C} p_c^{k_c}$

explanation: with replacement, unordered.

3.4 Binomial Distribution

Prerequisites: $n \in \mathbb{N}$, $p \in [0,1]$, $\Omega = \{0,\dots,n\}$
Name: $\mathbb{P} = \texttt{Bin}(n,p)$
Expectation: $\mathbb{E}(X) = np$
Variance: $\texttt{Var}(X) = np(1-p)$
MGF: $M_X(t) = q + pe^t$
pmf: $p_k := \binom{n}{k} p^k (1-p)^{n-k}$

explanation: discrete. with replacement. Look at a type (color) with probability p ("success rate", getting it in one try). Then, the Binomial Distribution is the probability of getting exactly k times this type out of n tries. A special case of this distribution is the Bernoulli Distribution for n=1.

3.5 Hypergeometric Distribution

Prerequisites: C finite set, $|C| \geq 2$, $(N_c)_{c\in C}$ in $\mathbb{N}^C$, $N := \sum_{c\in C} N_c$, $n \in \mathbb{N}$, $\Omega := \{(k_c)_{c\in C} \in \mathbb{N}_0^C; \sum_{c\in C} k_c = n\}$
Name: $\mathbb{P} = \texttt{HypGeo}(n,(N_c)_{c\in C})$
Expectation: $\mathbb{E}(X_c) = \frac{nN_c}{N}$
Variance: $\texttt{Var}(X_c) = \frac{N_c}{N}(1 - \frac{N_c}{N}) n \frac{N-n}{N-1}$

pmf: $p_{(k_c)_{c\in C}} := \frac{\prod_{c\in C}\binom{N_c}{k_c}}{\binom{N}{n}}$

explanation: without replacement, discrete, at least 2 types. Search for all balls with types (”colors”) in $(N_c)_{c\in C}$, N being the number of all balls interested in. Then, this gives the probability of getting exactly k_1 times balls of color 1, k_2 times balls of color 2 and so on, out of n times of drawing a ball without replacement.

3.6 Poisson Distribution

Prerequisites: $\Omega := \mathbb{N}_0, \lambda > 0$
Name: $\mathbb{P} = \texttt{Poi}(\lambda)$
Expectation: $\mathbb{E}(X) = \lambda$
Variance: $\texttt{Var}(X) = \lambda$
MGF: $M_X(t) = e^{\lambda(e^t-1)}$
pmf: $p_k = \frac{\lambda^k}{k!}e^{-\lambda}$

explanation: discrete. This is used to model the number of occurrences in a fixed time interval.

For questions which includes keywords ”at least g” use the sum from g to infinity.

3.7 Continuous Uniform Distribution

Prerequisites: $\Omega \in \mathcal{B}(\mathbb{R}^n)$ bounded with positive volume measure $\lambda^n(\Omega)$
Name: $\mathbb{P} = \texttt{Uniform}(\Omega)$
Expectation: $\mathbb{E}(X) = \frac{1}{2}(a+b)$ (for $\Omega = [a, b]$)
Variance: $\texttt{Var}(X) = \frac{1}{12}(b-a)^2$
MGF: $M_X(t) = \frac{e^{tb}-e^{ta}}{t(b-a)}$
pdf: $f(x) := \frac{1}{\lambda^n(\Omega)}$ ($= \frac{1}{b-a}$ for $\Omega = [a, b]$)

explanation: continuous. Every interval of the same size (within Ω) has the same probability to happen.

3.8 Waiting Times

Waiting times model the number of repetitions (discrete case) or the time (continuous case) until a certain event happens.

3.8.1 Negative Binomial Distribution

Prerequisites: $\Omega := \mathbb{N}_0$, $p \in (0, 1]$, $r \in \mathbb{N}$
Name: $\mathbb{P} = \texttt{NegBin}(r, p)$
Expectation: $\mathbb{E}(X) = \frac{r}{p} - r = \frac{r(1-p)}{p}$
Variance: $\texttt{Var}(X) = \frac{r(1-p)}{p^2}$
MGF: $M_X(t) = (\frac{p}{1-qe^t})^r$, $qe^t < 1$
pmf: $p_k := \binom{r+k-1}{k}p^r(1-p)^k = \binom{-r}{k}p^r(p-1)^k$

explanation: discrete. There are only 2 types and p is the probability of getting one of them. We draw from an urn until we receive r balls of this type. k is the number of balls not having this color and it is true that $r + k = n$, n being the number of draws.

3.8.2 Geometric Distribution

Prerequisites: $\Omega := \mathbb{N}_0$, $p \in (0,1]$
Name: $\mathbb{P} = \texttt{Geo}(p)$
Expectation: $\mathbb{E}(X) = \frac{1-p}{p}$
Variance: $\texttt{Var}(X) = \frac{1-p}{p^2} = \frac{1}{p^2} - \frac{1}{p}$
MGF: $M_X(t) = \frac{p}{1-qe^t}$, $qe^t < 1$
pmf: $p_k := (1-p)^k p$

explanation: discrete. The waiting time for the first occurrence, i.e. $r = 1$.

3.8.3 Gamma Distribution

Prerequisites: $\Omega := [0, \infty)$, $\alpha > 0$, $r > 0$
Name: $\mathbb{P} = \texttt{Gamma}(\alpha, r)$
Expectation: $\mathbb{E}(X) = \frac{r}{\alpha}$
Variance: $\texttt{Var}(X) = \frac{r}{\alpha^2}$
MGF: $M_X(t) = (1 - \frac{t}{r})^{-\alpha}$, $t < \alpha$
pdf: $f(x) := \frac{\alpha^r}{\Gamma(r)} x^{r-1} e^{-\alpha x}$

explanation: continuous. $\mathcal{A} := \mathcal{B}(\Omega)$ and $\Gamma(x) := \int\limits_0^\infty t^{x-1} e^{-t} dt$ and $\Gamma(n) = (n-1)!$ for $n \in \mathbb{N}$. Consider a closed interval A. This yields $\mathbb{P}(A) := \int\limits_A \frac{\alpha^r}{\Gamma(r)} x^{r-1} e^{-\alpha x} dx$ $(A \in \mathcal{A})$.

Remarks Gamma Function:

$\Gamma(x+1) = x \cdot \Gamma(x)$

$\Gamma(x) \cdot \Gamma(1-x) = \frac{\pi}{\sin x\pi}$

$\Gamma(x) \cdot \Gamma(x + \frac{1}{2}) = \frac{\sqrt{\pi}}{2^{2x-1}} \cdot \Gamma(2x)$

$\Gamma(\frac{1}{2}) = \sqrt{\pi}$

$\Gamma(-\frac{1}{2}) = -2\sqrt{\pi}$

$\Gamma(\frac{3}{2}) = \frac{1}{2}\sqrt{\pi}$

3.8.4 Exponential Distribution

Prerequisites: $\Omega := [0, \infty)$, $\lambda > 0$
Name: $\mathbb{P} = \texttt{Exp}(\lambda)$
Expectation: $\mathbb{E}(X) = \frac{1}{\lambda}$
Variance: $\texttt{Var}(X) = \frac{1}{\lambda^2}$
MGF: $M_X(t) = \frac{\lambda}{\lambda - t}$, $t < \lambda$
pdf: $f(x) := \lambda e^{-\lambda x}$

explanation: continuous. The Gamma distribution for $r = 1$, waiting for the first occurrence.

3.9 Normal Distribution

Prerequisites: $\Omega := \mathbb{R}$, $\mu \in \mathbb{R}$, $\sigma > 0$
Name: $\mathbb{P} = \texttt{Normal}(\mu, \sigma^2)$
Expectation: $\mathbb{E}(X) = \mu$
Variance: $\texttt{Var}(X) = \sigma^2$
MGF: $M_X(t) = e^{t\mu + \frac{\sigma^2 t^2}{2}}$
pdf: $f(x) := \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
explanation: Turning points are at $x = \mu \pm \sigma$.

3.10 Multidimensional Normal Distribution

Prerequisites: $\Omega := \mathbb{R}$ $\mu \in \mathbb{R}$, $\Sigma \in \mathbb{R}^{n \times n}$ symmetric and positive definite
Name: $\mathbb{P} = \texttt{Normal}(\mu, \Sigma)$
Expectation: $\mathbb{E}(X) = \mu$
pdf: $f_X(s) := (2\pi)^{\frac{n}{2}} \det(\Sigma)^{-\frac{1}{2}} e^{-\frac{1}{2}(s-\mu)^\top \Sigma^{-1}(s-\mu)}$

3.11 Distribution Properties

3.11.1 Important CDFs

Standard Normal Φ

Exponential(λ) $F(x) = 1 - e^{-\lambda x}, for x \in [0, \infty)$

Uniform(0,1) $F(x) = x, for x \in (0, 1)$

3.11.2 Convolution of RVs (X+Y)

$X \sim Poi(\lambda_1), Y \sim Poi(\lambda_2) \to X + Y \sim Poi(\lambda_1 + \lambda_2)$

$X \sim Bin(n_1, p), Y \sim Bin(n_2, p) \to X + Y \sim Bin(n_1 + n_2, p)$

$X \sim Gamma(a_1, \lambda), Y \sim Gamma(a_2, \lambda) \to X + Y \sim Gamma(a_1 + a_2, \lambda)$

$X \sim N.Bin(r_1, p), Y \sim N.Bin(r_2, p) \to X + Y \sim N.Bin(r_1 + r_2, p)$

$X \sim \texttt{Normal}(\mu_1, \sigma_1^2), Y \sim \texttt{Normal}(\mu_2, \sigma_2^2) \to X + Y \sim \texttt{Normal}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

3.11.3 Special Cases of Distributions

$\texttt{Bin}(1, p) \sim \texttt{Bernoulli}(p)$

$\texttt{Beta}(1, 1) \sim \texttt{Uniform}(0, 1)$

$\texttt{Gamma}(1, \lambda) \sim \texttt{Exp}(\lambda)$

$\texttt{X}_\texttt{n}^2() \sim \texttt{Gamma}(\frac{n}{2}, \frac{1}{2})$

$\texttt{NegBin}(1, p) \sim \texttt{Geo}(p)$

3.11.4 Inequalities

Cauchy-Schwarz $|E(XY)| \leq \sqrt{E(X^2)E(Y^2)}$

Markov $\mathbb{P}(X \geq a) \leq \frac{E(|X|)}{a}, (a > 0)$

Chebyshev $\mathbb{P}(|X - E(X) \geq a|) \leq \frac{Var(X)}{a^2}$ and

$\mathbb{P}(|X - E(X) \geq k \cdot \sigma|) \leq \frac{1}{k^2}$

4 Other Probabilities

4.1 Conditional Probabilities

$A \cap B$: region of both A and B, not just A or just B

$A \cup B$: both A and B

$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = c_B \mathbb{P}(A)$: the conditional probability of A given B

Bayes: Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $A, B \in \mathcal{A}$ with $\mathbb{P}(A), \mathbb{P}(B) > 0$. Then

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(B)\mathbb{P}(A \mid B)}{\mathbb{P}(A)}.$$

Total Probability, Case Distinction Formula: Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $n \in \mathbb{N}, B_1, ..., B_n \in \mathcal{A}$ pairwise disjoint such that $\Omega = \cup B_k (for 1 \leq k \leq n)$ and $\mathbb{P}(B_k) \geq 0$ for all $k \in \{1, ..., n\}$. Then

$$\mathbb{P}(A) = \sum_{k=1}^{n} \mathbb{P}(B_k)\mathbb{P}(A \mid B_k)(for A \in \mathcal{A}).$$

Multi-Stage Models: (2.37) Lemma) (Multiplication Rule). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $n \in \mathbb{N}, A_1, ..., A_n \in \mathcal{A}$. Then

$$\mathbb{P}(A_1 \cap ... \cap A_n) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2 \mid A_1) \cdot ... \cdot \mathbb{P}(A_n \mid A_1 \cap ... \cap A_{n-1})$$

4.2 Independence

Independence of two events: Independence is there, if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Independence of a family of events: Independence is there, if $\mathbb{P}(A \cap B \cap ... \cap Z) = \mathbb{P}(A)\mathbb{P}(B)...\mathbb{P}(Z)$. This is much stronger than say pairwise independence.

5 R

5.1 General Advice

In RStudio, you can type `?functionName` or `help(functionName)` into the interpreter/console in order to learn more about a function such as what arguments it takes and what they do. As such, for the sake of brevity, function names should be sufficient here as RStudio will also be available during the exam (I asked our professor beforehand).

5.2 Distributions

Add the prefix you want to the function stem to get the full function. ~~For example `runif(1, min = 0, max = 1)` will give you one random value according to a uniform distribution between 0 and 1.~~

Function Prefix	Meaning
d	density function
p	distribution function
q	quantile function
r	generate random value according to distribution

Distribution	Function Stem
Discrete Uniform	dunif
Bernoulli	bern
Multinomial	multinom
Binomial	binom
Hypergeometric	hyper
Poisson	pois
Continuous Uniform	unif
Negative Binomial	nbinom
Geometric	geom
Gamma	gamma
Exponential	exp
Normal	norm

5.3 Graphs

Function for creating histograms: `hist(...)`

Simple plotting: `plot(...)`

Declaring a color for graphs:

`histColor = rgb(red, green, blue, alpha, names = NULL, maxColorValue = 1)`

Add Connected Line Segments to a Plot: `lines(...)`

Set or Query Graphical Parameters: `par(...)`

5.4 Stochastics Functions

Variance: `var(x)`

Median: `median(x)`

Mean: `mean(x)`

For urn models: `sample(x, size, replace = FALSE, prob = NULL)`.

5.5 Math

pi: `pi`

Standard functions: `cos(); sin(); exp();`

5.6 Operators

```
a %% b  # a modulo b
a %/% b # integer division: a/b
```

```
a:b     # create an array of all integers from a to b (useful for loops)
```

5.7 Other useful functions

Replicating a function: `replicate(n, function())`

Declaring an array of numbers: `X = numeric(length)`

Creating an array or list: `X = c(value1, value2, ...)`

Creating a matrix: `matrix(...)`

Getting the length of an array or list: `length(X)`

Generating regular sequences: `seq(...)`

Summing up every element of an array: sum(...)

Combining vectors into matrices:

```
rbind(X1, X2, ...) # interprets X1, X2, ... as rows
cbind(X1, X2, ...) # interprets X1, X2, ... as columns
```

Concatenating objects after turning them into strings: `paste(..., sep = " ", collapse = NULL)`

Loading a CSV file: `read.csv(...)`

6 Calculus and Basic Functions

f $\rightarrow$ f'

ln(x) $\rightarrow \frac{1}{x}$

$a^x \rightarrow x^x(1 + ln(x))$

tan(x) $\rightarrow \frac{1}{cos^2(x)}$

cot(x) $\rightarrow \frac{-1}{sin^2(x)}$

arcsin(x) $\rightarrow \frac{1}{\sqrt{1-x^2}}$

arccos(x) $\rightarrow \frac{-1}{\sqrt{1-x^2}}$

arctan(x) $\rightarrow \frac{1}{1+x^2}$

arccot(x) $\rightarrow \frac{-1}{1+x^2}$

arcsinh(x) $\rightarrow \frac{1}{\sqrt{x^2+1}}$

arccosh(x) $\rightarrow \frac{1}{\sqrt{x^2-1}}$, x > 1

arctanh(x) $\rightarrow \frac{1}{1-x^2}$, $|x| < 1$

arccoth(x) $\rightarrow \frac{1}{1-x^2}$, $|x| > 1$

Natural Logarithm:

$ln(x \cdot z) = ln(x) + ln(z)$.

$ln(\frac{x}{z}) = ln(x) - ln(z)$.

$ln(x^r) = rln(x)$.

$ln(e^r) = r$.

$ln(e) = 1, ln(1) = 0$.

Integrals:

$\int \frac{f'}{f} dx = ln|f|$

$\int \frac{1}{x} dx = ln|x|$

$\int \frac{1}{x+a} dx = ln|x + a|$

$\int \frac{1}{(x+1)^2} dx = \frac{-1}{x+a}$

$\int \frac{1}{\sqrt{x}} dx = 2\sqrt{x}$

$\int ln|x|dx = xln(x) - x$

$\int ln(x)dx = x^2(\frac{ln(x)}{2} - \frac{1}{4})$

Differentiation Rules:

Product Rule: $(u \cdot v)' = u'v + uv'$

Quotient Rule: $(\frac{u}{v})' = \frac{u'v - uv'}{v^2}$

Chain rule: $u(v)' = u'(v) \cdot v'$

Integration Rules:

Integration by Parts: $\int_a^b u(x) \cdot v'(x)dx = u(x) \cdot v(x)\Big|_a^b - \int_a^b v(x) \cdot u'(x)dx$

Integration by Substitution: $\int_{\varphi(a)}^{\varphi(b)} f(u)du = \int_a^b f(\varphi(x)) \cdot \varphi'(x)dx, \quad du = \varphi'(x)dx$