FORMULAE SHEET
for
Probability and Statistics
FS 2021

1 Definitions

1. **Probability space**

   A probability space is a triple $(\Omega, \mathcal{A}, \mathbb{P})$, where:

   (a) Ω is an arbitrary non-empty set.

   (b) $\mathcal{A}$ is a σ-algebra on Ω:

      i. $\Omega \in \mathcal{A}$

      ii. $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$

      iii. $A_1, A_2, \ldots \in \mathcal{A} \Rightarrow \cup_{n=1}^{\infty} A_n \in \mathcal{A}$

   (c) $\mathbb{P}\colon \mathcal{A} \to [0,1]$ is a probability measure on $(\Omega, \mathcal{A})$:

      i. $P(\Omega) = 1$

      ii. For each sequence $A_1, A_2, \ldots \in \mathcal{A}$ such that $\forall i \neq j\colon A_i \cap A_j = \emptyset$, it holds that $\mathbb{P}(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$

2. **Random variable**

   A random variable X is a $\mathcal{B}(\mathbb{R})/\mathcal{A}$-measurable map $X\colon \Omega \to \mathbb{R}$.

3. **Distribution and distribution function**

   The distribution of a random variable X is the probability measure $\mu_X\colon \mathcal{B}(\mathbb{R}) \to [0,1]$ satisfying

   $$\forall\, B \in \mathcal{B}(\mathbb{R})\colon \quad \mu_X(B) = \mathbb{P}(X \in B).$$

   The cumulative distribution function (CDF) of X is the function $F_X\colon \mathbb{R} \to [0,1]$ defined by

   $$\forall\, x \in \mathbb{R}\colon \quad F_X(x) = \mathbb{P}(X \leq x) = \mu_X((-\infty, x]).$$

4. **Expectation**

   The expected value of a random variable X, if it exists, is given by

   $$\mathbb{E}[X] = \int_\Omega X(\omega)\mathbb{P}(d\omega) = \int_\mathbb{R} x\mu_X(dx) = \begin{cases} \sum_x x\mathbb{P}(X = x), & \text{discrete case} \\ \int_\mathbb{R} x f_X(x)dx, & \text{absolutely continuous case} \end{cases}.$$

5. **Variance**

   The variance of a random variable X with $\mathbb{E}[|X|] < \infty$ is given by

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

6. **Covariance**

   The covariance of two random variables X, Y with $\mathbb{E}[|X|] < \infty$, $\mathbb{E}[|Y|] < \infty$, $\mathbb{E}[|XY|] < \infty$ is defined by

   $$\begin{aligned}\mathrm{Cov}(X,Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].\end{aligned}$$

7. **Correlation**

   The correlation of two random variables X, Y with $0 < \mathrm{Var}(X) < \infty$, $0 < \mathrm{Var}(Y) < \infty$ is defined by

   $$\mathrm{corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}.$$

8. **Moment-generating function**

   The moment-generating function of a real-valued random variable X is defined as

   $$\begin{aligned} M_X\colon \mathbb{R} &\to [0,\infty] \\ t &\mapsto M_X(t) = \mathbb{E}[e^{tX}]. \end{aligned}$$

9. **Characteristic function**

   The characteristic function of a real-valued random variable X is defined as

   $$\begin{aligned} \varphi_X\colon \mathbb{R} &\to \mathbb{C} \\ t &\mapsto \varphi_X(t) = \mathbb{E}[e^{itX}]. \end{aligned}$$

2 Some discrete distributions

1. **Bernoulli distribution**

   $X \sim \mathrm{Ber}(p)$ for $p \in [0,1]$:

   $$\mathbb{P}(X = x) = \begin{cases} p, & x = 1 \\ 1-p, & x = 0 \end{cases}$$

   $\mathbb{E}[X] = p$, $\mathrm{Var}(X) = p(1-p)$.

2. **Binomial distribution**

   $X \sim \mathrm{Bin}(n,p)$ for $n \in \mathbb{N}$, $p \in [0,1]$:

   $$\forall x \in \{0, 1, \ldots, n\}: \quad \mathbb{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

   $\mathbb{E}[X] = np$, $\mathrm{Var}(X) = np(1-p)$.

3. **Geometric distribution**

   $X \sim \mathrm{Geo}(p)$ for $p \in (0,1)$:

   $$\forall x \in \mathbb{N}: \quad \mathbb{P}(X = x) = (1-p)^{x-1}p$$

   $\mathbb{E}[X] = \frac{1}{p}$, $\mathrm{Var}(X) = \frac{1-p}{p^2}$.

4. **Poisson distribution**

   $X \sim \mathrm{Poi}(\lambda)$ for $\lambda > 0$:

   $$\forall x \in \mathbb{N}_0: \quad \mathbb{P}(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

   $\mathbb{E}[X] = \lambda$, $\mathrm{Var}(X) = \lambda$.

5. **Negative binomial distribution**

   $X \sim \mathrm{NB}(r,p)$ for $p \in (0,1), r \in \mathbb{N}$:

$$\forall x \in \{0,1,2,\ldots\}: \quad \mathbb{P}(X=x) = \binom{x+r-1}{x} p^r (1-p)^x$$

   $\mathbb{E}[X] = \frac{r(1-p)}{p}$, $\mathrm{Var}(X) = \frac{r(1-p)}{p^2}$.

6. **Hypergeometric distribution**

   $X \sim \mathrm{Hypergeom}(n,N,K)$ for $n,N,K \in \mathbb{N}$ such that $\max(n,K) \leq N$:

$$\forall x \in \{\max(0, n-N+K), ..., \min(n,K)\}: \quad \mathbb{P}(X=x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$$

   $\mathbb{E}[X] = n\frac{K}{N}$, $\mathrm{Var}(X) = n\frac{K}{N}\left(1-\frac{K}{N}\right)\frac{N-n}{N-1}$.

3 Some continuous distributions

1. **Uniform distribution**

   $X \sim \mathcal{U}(a,b)$ for $a<b$:

$$\forall x \in \mathbb{R}: \quad f_X(x) = \frac{1}{b-a}\mathbb{1}_{[a,b]}(x)$$

   $\mathbb{E}[X] = \frac{a+b}{2}$, $\mathrm{Var}(X) = \frac{(b-a)^2}{12}$.

2. **Exponential distribution**

   $X \sim \mathrm{Exp}(\lambda)$ for $\lambda > 0$:

$$\forall x \in \mathbb{R}: \quad f_X(x) = \lambda e^{-\lambda x}\mathbb{1}_{(0,\infty)}(x)$$

   $\mathbb{E}[X] = \frac{1}{\lambda}$, $\mathrm{Var}(X) = \frac{1}{\lambda^2}$.

3. **Gamma distribution**

   $X \sim \Gamma(\alpha,\beta)$ for $\alpha, \beta > 0$:

$$\forall x \in \mathbb{R}: \quad f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}\mathbb{1}_{(0,\infty)}(x),$$

   $\mathbb{E}[X] = \frac{\alpha}{\beta}$, $\mathrm{Var}(X) = \frac{\alpha}{\beta^2}$.

4. **Normal distribution**

   $X \sim \mathcal{N}(\mu,\sigma^2)$ for $\mu \in \mathbb{R}, \sigma > 0$:

$$\forall x \in \mathbb{R}: \quad f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

   $\mathbb{E}[X] = \mu$, $\mathrm{Var}(X) = \sigma^2$.

   If $\mu = 0$ and $\sigma = 1$, we say that X is standard normal and its distribution function (CDF) is denoted by Φ.

4 Independence and Bayes' theorem

1. **Independence**

   Let I be an arbitrary set and $(A_i)_{i \in I} \subseteq \mathcal{A}$ a family of events. Then, $(A_i)_{i\in I}$ is said to be independent if

$$\forall J \subseteq I \text{ finite}: \quad \mathbb{P}(\cap_{j \in J} A_j) = \prod_{j \in J} \mathbb{P}(A_j).$$

2. **Conditional probability**

   Let $A, B \in \mathcal{A}$ be two events with $\mathbb{P}(B) > 0$. The conditional probability of A given B is

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

3. **Law of total probability**

   Let $(B_i)_{i \in I} \subseteq \mathcal{A}$ be a family of events such that $\Omega = \cup_{i \in I} B_i$ and $\forall\, i \neq j\colon B_i \cap B_j = \emptyset$. Then,

$$\forall A \in \mathcal{A}: \quad \mathbb{P}(A) = \sum_{i \in I:\, \mathbb{P}(B_i)>0} \mathbb{P}(A|B_i)\mathbb{P}(B_i).$$

4. **Bayes' theorem**

   Let $A, B \in \mathcal{A}$ be two events with $\mathbb{P}(A) > 0$. Then,

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c)}.$$

5 Borel-Cantelli

Let $A_1, A_2, \ldots \in \mathcal{A}$ be a sequence of events and set

$$A_\infty = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m.$$

1. If $\sum_{n=1}^\infty \mathbb{P}(A_n) < \infty$, then $\mathbb{P}(A_\infty) = 0$.
2. If the events $(A_n)_{n=1}^\infty$ are mutually independent and satisfy $\sum_{n=1}^\infty \mathbb{P}(A_n) = \infty$, then $\mathbb{P}(A_\infty) = 1$.

6 Limit Theorems

1. **Convergence of random variables**

   (a) **Convergence in probability**

   Let $X, X_1, X_2, \ldots$ be random variables. We say that the sequence $(X_n)_{n=1}^\infty$ converges in probability to X if

$$\forall\, \varepsilon > 0: \quad \lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0.$$

   (b) **Almost sure convergence**

   Let $X, X_1, X_2, \ldots$ be random variables. We say that the sequence $(X_n)_{n=1}^\infty$ converges $\mathbb{P}$-almost surely ($\mathbb{P}$-a.s.) to X if

$$\mathbb{P}(\{\omega \in \Omega\colon \lim_{n\to\infty} |X_n(\omega) - X(\omega)| = 0\}) = 1.$$

   (c) **Convergence in distribution**

   Let $\mu, \mu_1, \mu_2, \ldots$ be distributions over $\mathbb{R}$ and let $F, F_1, F_2, \ldots$ be the corresponding distribution functions. We say that $(\mu_n)_{n=1}^\infty$ converges weakly (in distribution) to μ if one of the following two equivalent statements is true:

   i. $\forall\, f \in C_b(\mathbb{R})\colon \lim_{n\to\infty} \int_{\mathbb{R}} f d\mu_n = \int_{\mathbb{R}} f d\mu$.

   ii. $\forall\, x \in \mathbb{R}$, F cont. at x: $\lim_{n\to\infty} F_n(x) = F(x)$.

   (d) **Properties**

   Almost sure convergence $\Rightarrow$ convergence in probability $\Rightarrow$ convergence in distribution.

2. **Strong law of large numbers**

   Let $X_1, X_2, \ldots$ be an i.i.d. sequence of random variables with $\mathbb{E}[|X_1|^2] < \infty$. Set $S_n = \sum_{i=1}^n X_i$ for any $n \in \mathbb{N}$. Then, S_n/n converges almost surely to $\mathbb{E}[X_1]$.

3. **Central limit theorem**

   Let $X_1, X_2, \ldots$ be an i.i.d. sequence of random variables with $\mathbb{E}[|X_1|^2] < \infty$. Set $\mu = \mathbb{E}[X_1]$ and $\sigma^2 = \mathrm{Var}(X_1)$. Let $S_n^* = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}}$ for any $n \in \mathbb{N}$. Then, the distribution of S_n^* converges to the standard normal distribution $\mathcal{N}(0,1)$:

$$\forall x \in \mathbb{R}: \quad \lim_{n\to\infty} \mathbb{P}(S_n^* \leq z) = \Phi(z).$$

7 Some inequalities

1. **Markov's inequality**

   Let X be a random variable and $g\colon \mathbb{R} \to [0,\infty)$ increasing. For any c with $g(c) > 0$,

$$\mathbb{P}(X \geq c) \leq \frac{\mathbb{E}[g(X)]}{g(c)}.$$

2. **Chebyshev's inequality**

   Let X be a random variable with $\mathbb{E}[|X|^2] < \infty$. For any $c > 0$,

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

3. **Cauchy–Schwarz inequality**

   Let X, Y be random variables with $\mathbb{E}[X^2] < \infty$, $\mathbb{E}[Y^2] < \infty$. Then,

$$|\mathbb{E}[XY]| \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}.$$

4. **Jensen's inequality**

   Let X be a random variable with $\mathbb{E}[|X|] < \infty$ and $g\colon \mathbb{R} \to \mathbb{R}$ a convex function. Then,

$$\mathbb{E}[g(X)] \geq g(\mathbb{E}[X]).$$

   If g is strictly convex, then equality holds if and only if $\mathbb{P}(X = \mathbb{E}[X]) = 1$.

8 Point estimators

Let $X = (X_1, X_2, \ldots, X_n)$ be a $\mathbb{R}^n$-valued random vector. Let Θ be a set and let $(\mu_\theta)_{\theta\in\Theta}$ be a family of possible distributions of X. Let $g\colon \theta \to \mathbb{R}^k$ be a map. We are interested in estimating a parameter of interest $\eta = g(\theta)$ using X.

1. **Definition**

   A point estimator of $g(\theta)$ is a map

$$T\colon \mathbb{R}^n \to \mathbb{R}^k.$$

2. **Unbiased**

   Let T be a point estimator of $g(\theta)$. The estimator T is called unbiased for $g(\theta)$ if

$$\forall\, \theta \in \Theta: \quad \mathbb{E}_\theta[T(X)] = g(\theta).$$

3. **Consistency**

   Assume n to be a variable and consider a sequence $T_n: \mathbb{R}^n \to \mathbb{R}$, $n \in \mathbb{N}$, of point estimators of $g(\theta) \in \mathbb{R}$. The sequence $(T_n)_{n\in\mathbb{N}}$ is called consistent for $g(\theta)$ if

$$\forall\, \theta \in \Theta, \varepsilon > 0\colon \ \limsup_{n\to\infty} \mathbb{P}_\theta(|T_n(X_1, \ldots, X_n) - g(\theta)| > \varepsilon) = 0.$$

9 Statistical hypothesis testing

We consider a null hypothesis $\Theta_0 \subseteq \Theta$ and an alternative hypothesis $\Theta_A \subseteq \Theta_0^c$.

1. **Definition**

   A statistical test is a map $\varphi\colon \mathbb{R}^n \to \{0,1\}$. For $x \in \mathbb{R}^n$, we interpret:

   - $\varphi(x) = 0 \Rightarrow$ "The null hypothesis is kept"
   - $\varphi(x) = 1 \Rightarrow$ "The null hypothesis is rejected"

   A randomised test is a map $\varphi\colon \mathbb{R}^n \to [0,1]$. For $x \in \mathbb{R}^n$, we interpret $\varphi(x)$ as the probability of rejecting the null hypothesis and $(1 - \varphi(x))$ as the probability of keeping the null hypothesis.

2. **Type I and type II errors**

   - Type I error: The null hypothesis is rejected even though it is true.
   - Type II error: The null hypothesis is kept even though it is false.

3. **Level and power of a test**

   Let $\alpha \in [0,1]$. A test φ has level α for the null hypothesis Θ_0 if

$$\sup_{\theta\in\Theta_0} \mathbb{E}_\theta[\varphi(X)] \leq \alpha.$$

   For $\theta \in \Theta_0^c$, we call $\mathbb{E}_\theta[\varphi(X)]$ the power of the test φ.

4. **Neyman-Pearson lemma**

   Let $\Theta = \{0,1\}$, $\Theta_0 = \{0\}$, $\Theta_A = \{1\}$, let $p_i: \mathbb{R}^n \to [0,\infty)$ be the density of μ_i with respect to $\mu_0 + \mu_1$ for $i \in \{0,1\}$, and let $\alpha \in [0,1]$. Then,

   (a) there exists a randomised test φ and some $c \in [0,\infty]$ such that

$$\mathbb{E}_0[\varphi] = \alpha, \tag{1}$$

$$\forall\, x \in \mathbb{R}^n: \quad \varphi(x) = \begin{cases} 1 & \text{if } p_1(x) > cp_0(x) \\ 0 & \text{if } p_1(x) < cp_0(x) \end{cases}, \tag{2}$$

   (b) any test that satisfies (1) and (2) is a most powerful test at level α, and

   (c) any most powerful test at level α satisfies (2) $(\mu_0 + \mu_1)$-a.e. (almost everywhere). It also satisfies (1) except when there is a test φ' with $\mathbb{E}_1[\varphi'] = 1$ and $\mathbb{E}_0[\varphi'] < \alpha$.

   Remark:

   If μ_0 and μ_1 are absolutely continuous distributions with densities f_0 and f_1, respectively, then p_i can be replaced by f_i in point (a) of the lemma.

10 Some test statistics

Let $X_1, \ldots, X_n$ be i.i.d. random variables with an appropriate distribution (depending on the desired test). Let $Y_1, \ldots, Y_m$ be i.i.d. random variables which are independent of $X_1, \ldots, X_n$ and with an appropriate distribution (depending on the desired test). Set

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2,$$
$$\overline{Y}_m = \frac{1}{m}\sum_{k=1}^{m} Y_k, \quad S_Y^2 = \frac{1}{m-1}\sum_{k=1}^{m}(Y_k - \overline{Y}_m)^2.$$

1. **One-sample t-test**

$$T = \frac{\overline{X}_n - \mu_0}{S_X/\sqrt{n}}.$$

2. **z-test**

$$T = \frac{\overline{X}_n - \mu_0}{\sigma/\sqrt{n}}.$$

3. **Two-sample t-test**

$$T = \frac{\overline{X}_n - \overline{Y}_m}{\sqrt{\frac{1}{n}+\frac{1}{m}}\sqrt{\frac{1}{n+m-2}((n-1)S_X^2 + (m-1)S_Y^2}}.$$

4. **Sign test**

$$T = \sum_{i=1}^{n} \mathbb{1}_{\{X_i > \mu_0\}}.$$

5. **Two-sample Wilcoxon zesz**

$$T = \sum_{i=1}^{n}\sum_{k=1}^{m} \mathbb{1}_{\{X_i < Y_k\}}.$$

6. **Goodness-of-fit χ^2-test**

$$T = \sum_{i=1}^{k} \frac{(N_i - n\theta_{0,i})^2}{n\theta_{0,i}}.$$

In all cases, the distribution of the test statistic T under the null hypothesis is known (exactly or approximately), and the relevant quantiles can be found in the corresponding tables.