FORMULAE SHEET
for
Probability and Statistics
Spring 2019

1 Standard discrete distributions

1. **Bernoulli distribution**

   $X \sim \mathrm{Ber}(p)$ ($p \in (0,1)$). Then, the pmf is given by
   $$p_X(x) = \begin{cases} p, & x = 1 \\ 1-p, & x = 0 \end{cases} = p^x(1-p)^{1-x}.$$
   X satisfies: $E(X) = p$, $\mathrm{var}(X) = p(1-p)$.

2. **Binomial distribution**

   $X \sim \mathrm{Bin}(n,p)$ ($n \in \mathbb{N}$, $p \in (0,1)$). Then, the pmf is given by
   $$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x \in \{0, 1, ..., n\}.$$
   X satisfies: $E(X) = np$, $\mathrm{var}(X) = np(1-p)$.

3. **Geometric distribution**

   $X \sim \mathrm{Geo}(p)$ ($p \in (0,1)$). Then, the pmf is given by
   $$p_X(x) = (1-p)^{x-1}p, \quad x \in \{1, 2, ...\}.$$
   X satisfies: $E(X) = \frac{1}{p}$, $\mathrm{var}(X) = \frac{1-p}{p^2}$.

4. **Poisson distribution**

   $X \sim \mathrm{Poi}(\lambda)$ ($\lambda > 0$). Then, the pmf is given by
   $$p_X(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x \in \{0, 1, 2, ...\}.$$
   X satisfies: $E(X) = \lambda$, $\mathrm{var}(X) = \lambda$.

5. **Discrete uniform distribution**

   $X \sim \mathrm{Unif}\{a, a+1, ..., b\}$ ($a \leq b$ integers). Then, the pmf is given by
   $$p_X(x) = \frac{1}{b-a+1} \quad x \in \{a, a+1, ..., b\}.$$
   X satisfies: $E(X) = \frac{a+b}{2}$, $\mathrm{var}(X) = \frac{(b-a+1)^2-1}{12}$.
   For $a = 1$ and $b = N \geq 1$ an integer: $E(X) = \frac{N+1}{2}$, $\mathrm{var}(X) = \frac{N^2-1}{12}$.

6. **Negative binomial distribution**

   $X \sim \mathrm{NB}(r,p)$ ($p \in (0,1), r \in \{1, 2, ...\}$). Then, the pmf is given by
   $$p_X(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \quad x \in \{r, r+1, ...\}.$$
   X satisfies: $E(X) = \frac{r}{p}$, $\mathrm{var}(X) = \frac{r(1-p)}{p^2}$.

7. **Hypergeometric distribution**

   $X \sim \mathrm{Hypergeo}(n, D, N)$ (n, D, N positive integers such that $n, D \leq N$). Then, the pmf is given by
   $$p_X(x) = \frac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}}, \quad x \in \{\max(0, n-N+D), ..., \min(n, D)\}.$$
   X satisfies: $E(X) = n\frac{D}{N}$, $\mathrm{var}(X) = n\frac{D}{N}\left(1 - \frac{D}{N}\right)\left(\frac{N-n}{N-1}\right)$.

2 Standard continuous distributions

1. **Uniform distribution**

   $X \sim \mathrm{U}([a,b])$ ($a < b$). Then, the pdf is given by
   $$f_X(x) = \frac{1}{b-a} \mathbb{1}_{x \in [a,b]}.$$
   X satisfies: $E(X) = \frac{a+b}{2}$, $\mathrm{var}(X) = \frac{(b-a)^2}{12}$.

2. **Beta distribution**

   $X \sim \mathrm{B}(\alpha, \beta)$ ($\alpha, \beta > 0$). Then, the pdf is given by
   $$f_X(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \mathbb{1}_{x \in (0,1)}$$
   with
   $$\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt, \quad \text{for } a > 0.$$
   X satisfies: $E(X) = \frac{\alpha}{\alpha+\beta}$, $\mathrm{var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

3. **Exponential distribution**

   $X \sim \mathrm{Exp}(\lambda)$ ($\lambda > 0$). Then, the pdf is given by
   $$f_X(x) = \lambda e^{-\lambda x} \mathbb{1}_{x > 0}.$$
   X satisfies: $E(X) = \frac{1}{\lambda}$, $\mathrm{var}(X) = \frac{1}{\lambda^2}$.

4. **Gamma distribution**

   $X \sim \mathrm{G}(\alpha, \lambda)$ ($\alpha, \lambda > 0$). Then, the pdf is given by
   $$f_X(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \mathbb{1}_{x > 0}.$$
   X satisfies: $E(X) = \frac{\alpha}{\lambda}$, $\mathrm{var}(X) = \frac{\alpha}{\lambda^2}$.

5. **Normal distribution**

   $X \sim \mathcal{N}(\mu, \sigma^2)$ ($\mu \in \mathbb{R}, \sigma > 0$). Then, the pdf is given by
   $$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$
   X satisfies: $E(X) = \mu$, $\mathrm{var}(X) = \sigma^2$, $\Psi_X(t) = E[e^{tX}] = \exp\left(\mu t + \frac{\sigma^2}{2}t^2\right)$ for $t \in \mathbb{R}$.

   If $\mu = 0$ and $\sigma = 1$, we say that X is standard normal and its cdf is denoted by Φ.

   Here are some special values of the quantiles for the standard normal distribution:
   $$\begin{aligned} z_{0.95} &= \Phi^{-1}(0.95) \approx 1.65, \\ z_{0.975} &= \Phi^{-1}(0.975) \approx 1.96, \\ z_{0.99} &= \Phi^{-1}(0.99) \approx 2.33, \\ z_{0.995} &= \Phi^{-1}(0.995) \approx 2.57. \end{aligned}$$

6. **Multivariate normal distribution**

   $\mathbf{X} = (X_1, ..., X_n)^T \sim \mathcal{N}(\mu, \Sigma)$ for some $\mu \in \mathbb{R}^n$ and $\Sigma \in \mathbb{R}^{n\times n}$ symmetric and positive semi-definite.

   $\mathbf{X}$ satisfies $E(\mathbf{X}) = \mu$ with $\mu_i = E(X_i)$ $(1 \le i \le n)$ and Σ is the covariance matrix with $\Sigma_{ij} = \text{cov}(X_i, X_j)$ $(1 \le i, j \le n)$.

   If Σ is invertible (positive definite), then $\mathbf{X}$ admits a density with respect to Lebesgue measure on $\mathbb{R}^n$, given by

   $$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{|\det(\Sigma)|}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)}, \quad \mathbf{x} \in \mathbb{R}^n.$$

3 Some inequalities

1. **Markov's inequality**

   Let X be a non-negative random variable, $E(X) < \infty$ and $a > 0$. Then
   $$P(X \ge a) \le \frac{E(X)}{a}.$$

2. **Chebyshev's inequality**

   Let X be a random variable with $E(X) = \mu < \infty$ and $\text{var}(X) = \sigma^2 < \infty$, and let $a > 0$. Then
   $$P(|X - \mu| \ge a) \le \frac{\sigma^2}{a^2}.$$

3. **Cauchy-Schwarz inequality**

   Let X, Y be random variables defined on the same probability space with $E(X^2), E(Y^2) < \infty$. Then
   $$|E(XY)| \le \sqrt{E(X^2)E(Y^2)}$$
   with equality if and only if either $P(X = 0) = 1$ or there exists $a \in \mathbb{R}$ such that $P(Y = aX) = 1$.

4. **Minkowski's inequality**

   Let X, Y be random variables defined on the same probability space with $E(|X|^p), E(|Y|^p) < \infty$ for some $p \ge 1$. Then
   $$E(|X + Y|^p)^{\frac{1}{p}} \le E(|X|^p)^{\frac{1}{p}} + E(|Y|^p)^{\frac{1}{p}}.$$

5. **Jensen's inequality**

   Let $f : \mathbb{R} \to \mathbb{R}$ be a convex function and X a random variable such that $E(|X|) < \infty$ and $E(|f(X)|) < \infty$. Then
   $$E(f(X)) \ge f(E(X)).$$
   If f is strictly convex, then equality holds if and only if $P(X = E(X)) = 1$.

4 Miscellaneous

1. **Covariance**

   For X, Y two random variables defined on the same probability space with $E(|X|), E(|Y|), E(|XY|) < \infty$,
   $$\begin{aligned}\text{cov}(X, Y) &= E[(X - E(X))(Y - E(Y))] \\ &= E(XY) - E(X)E(Y).\end{aligned}$$
   For any $n, m \in \mathbb{N}$, $a_1, \ldots, a_n$, $b_1, \ldots, b_m$ real numbers, $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ random variables defined on the same probability space with well-defined $\text{cov}(X_i, Y_j)$ for $1 \le i \le n$ and $1 \le j \le m$ we have
   $$\text{cov}\left(\sum_{i=1}^n a_i X_i, \sum_{j=1}^m b_j Y_j\right) = \sum_{j=1}^m \sum_{i=1}^n a_i b_j \text{cov}(X_i, Y_j).$$

2. **Correlation**

   If X, Y are two random variables defined on the same probability space with $0 < \text{var}(X), \text{var}(Y) < \infty$, then
   $$\text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}.$$
   It always holds that $|\text{corr}(X, Y)| \le 1$. Equality occurs if and only if there exists $a \ne 0$ and $b \in \mathbb{R}$ such that $P(Y = aX + b) = 1$, with $\text{sgn}(a) = \text{sgn}(\text{corr}(X, Y))$.

3. **Mean of a random vector**

   If $\mathbf{X} = (X_1, ..., X_n)^T$ is a random vector in $\mathbb{R}^n$, such that $E(|X_i|) < \infty$ for $i = 1, \ldots, n$, then $E(\mathbf{X}) = \mu \in \mathbb{R}^n$ with $\mu_i = E(X_i)$ for $i = 1, ..., n$. For any vector $a \in \mathbb{R}^n$, $E(a^T\mathbf{X}) = a^T\mu$.

4. **Covariance matrix of a random vector**

   Let $\mathbf{X} = (X_1, ..., X_n)^T$ be a random vector in $\mathbb{R}^n$, such that $\text{cov}(X_i, X_j)$ is defined for $1 \le i, j \le n$. Then $\mathbf{X}$ admits a covariance matrix, $\Sigma \in \mathbb{R}^{n\times n}$, given by $\Sigma_{ij} = \text{cov}(X_i, X_j)$ for $1 \le i, j \le n$.

   The covariance matrix Σ is also given by
   $$\Sigma = E[(\mathbf{X} - \mu)(\mathbf{X} - \mu)^T]$$
   where $\mu = E(\mathbf{X})$.

   For any vector $a \in \mathbb{R}^n$, $\text{var}(a^T\mathbf{X}) = a^T\Sigma\, a$.

5. **Uncorrelation and independence**

   Let $\mathbf{X} = (X_1, ..., X_n)^T$ be a random vector in $\mathbb{R}^n$ with a well-defined covariance matrix Σ. If $X_1, \ldots, X_n$ are independent, then $\text{cov}(X_iX_j) = 0$ for all $i \ne j$. If $\text{var}(X_i) > 0$ for all $i = 1, \ldots, n$, then this is equivalent to $\text{corr}(X_i, X_j) = 0$ for all $i \ne j$. In this case, Σ is a diagonal matrix.

   In the special case where $\mathbf{X} = (X_1, ..., X_n)^T \sim \mathcal{N}(\mu, \Sigma)$, Σ is diagonal if and only if $X_1, \ldots, X_n$ are independent.

6. **Normal approximation for Poisson**

   Let $X = X_\lambda \sim \text{Poi}(\lambda)$, with $\lambda \in (0, \infty)$. Then,
   $$\frac{X_\lambda - \lambda}{\sqrt{\lambda}} \xrightarrow{d} \mathcal{N}(0, 1)$$
   as $\lambda \to \infty$.

7. **Fisher information**

   Consider a parametric model $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ where Θ is a parameter space and P_θ a probability measure defined (on a measurable space $(\mathcal{X}, \mathcal{B})$) admitting a density $f(\cdot|\theta)$ with respect to some σ-finite dominating measure μ on $(\mathcal{X}, \mathcal{B})$.

   Based on $X_1, \ldots, X_n \in \mathcal{X}$ i.i.d. random variables $\sim f(\cdot|\theta_0)$ for some unknown $\theta_0 \in \Theta$, we want to estimate θ_0.

   Under some regularity conditions, the Fisher information of the model at θ_0 is given by
   $$\begin{aligned} I(\theta_0) &= E\left[\left(\frac{\partial \log f(X_1|\theta)}{\partial\theta}\Big|_{\theta=\theta_0}\right)^2\right] \\ &= -E\left[\frac{\partial^2 \log f(X_1|\theta)}{\partial^2\theta}\Big|_{\theta=\theta_0}\right].\end{aligned}$$