PROBABILITY THEORY

Sem. 1, Euler's Functions

Euler's Gamma Function $\Gamma:(0,\infty)\to(0,\infty)\quad \Gamma(a)=\int\limits_0^\infty x^{a-1}e^{-x}dx$

1. $\Gamma(1)=1$; **2.** $\Gamma(a+1)=a\Gamma(a)$, $\forall a>0$;

3. $\Gamma(n+1)=n!$, $\forall n\in\mathbb{N}$; **4.** $\Gamma\left(\frac{1}{2}\right)=\sqrt{2}\int\limits_0^\infty e^{-\frac{t^2}{2}}dt=\int\limits_{\mathbb{R}} e^{-t^2}dt=\sqrt{\pi}$.

Euler's Beta Function $\beta:(0,\infty)\times(0,\infty)\to(0,\infty)\quad \beta(a,b)=\int\limits_0^1 x^{a-1}(1-x)^{b-1}dx$

1. $\beta(a,1)=\frac{1}{a}$, $\forall a>0$; **2.** $\beta(a,b)=\beta(b,a)$, $\forall a,b>0$; **3.** $\beta(a,b)=\frac{a-1}{b}\beta(a-1,b+1)$, $\forall a>1,b>0$;
4. $\beta(a,b)=\frac{b-1}{a+b-1}\beta(a,b-1)=\frac{a-1}{a+b-1}\beta(a-1,b)$, $\forall a,b>1$; **5.** $\beta(a,b)=\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$, $\forall a,b>0$.

Sem. 2, Class. Prob., Geom. Prob., Cond. Prob., Indep. Events, Bayes' Formula

Classical Probability: $P(A)=\dfrac{\text{nr. of favorable outcomes}}{\text{total nr. of possible outcomes}}$.

Conditional Probability: $P(A|B)=\dfrac{P(A\cap B)}{P(B)}$, $P(B)\neq 0$.

Independent Events: A, B independent $<=>$ $P(A\cap B)=P(A)P(B)$ $<=>$ $P(A|B)=P(A)$

Total Probability Rule: $\{A_i\}_{i\in I}$ a partition of S, then $P(A)=\sum\limits_{i\in I}P(A_i)P(A_i|A)$

Multiplication Rule: $P\left(\bigcap\limits_{i=1}^{n}\right)A_i=P\left(A_1\right)P\left(A_2|A_1\right)P\left(A_3|A_1\cap A_2\right)\ ...\ P\left(A_n|\bigcap\limits_{i=1}^{n-1}A_i\right)$

Bayes' Formula: $\{A_i\}_{i\in I}$ a partition of S, then $P\left(A_j|A\right)=\dfrac{P\left(A|A_j\right)P\left(A_j\right)}{\sum\limits_{i\in I}P\left(A|A_i\right)P\left(A_i\right)}$, $\forall j\in I$

Sem. 3, Probabilistic Models

Binomial Model: The probability of k successes in n Bernoulli trials, with probability of success p, is $P(n,k)=C_n^k p^k q^{n-k},\ k=\overline{0,n}$.

Multinomial Model: The probability that in $n=n_1+n_2+...+n_r$ trials, E_i occurs n_i times, where $p_i=P\left(E_i\right)$, $i=\overline{1,r}$, is $P(n;n_1,...,n_r)=\dfrac{n!}{n_1!n_2!...n_r!}p_1^{n_1}p_2^{n_2}...p_r^{n_r}$.

Bernoulli Model Without Replacement (Hypergeometric): The probability that in n trials, we get k white balls out of n_1 and $n-k$ black balls out of $N-n_1$ $(0\le k\le n_1,\ 0\le n-k\le N-n_1)$, is $P(n;k)=\dfrac{C_{n_1}^k C_{N-n_1}^{n-k}}{C_N^n}$.

Bernoulli Model Without Replacement With r States: The probability that in $M=m_1+m_2+...+m_r$ trials, we get m_i balls of color i out of n_i, $i=\overline{1,r}$, $(n=n_1+n_2+...+n_r)$, is $P(n;m_1,...,m_r)=\dfrac{C_{n_1}^{m_1}C_{n_2}^{m_2}...C_{n_r}^{m_r}}{C_n^M}$.

Poisson Model: The probability of k successes $(0\le k\le n)$ in n trials, with probability of success p_i in the i^{th} trial $(q_i=1-p_i)$, $i=\overline{1,n}$, is $P(n;k)=\sum\limits_{1\le i_1<...<i_k\le n}p_{i_1}...p_{i_k}q_{i_{k+1}}...q_{i_n},\quad i_{k+1},...,i_n\in\{1,...,n\}\backslash\{i_1,...,i_k\}$
$=$ the coefficient of x^k in the expansion $(p_1x+q_1)(p_2x+q_2)\dots(p_nx+q_n)$.

Pascal Model: The probability of the n^{th} success occurring after k failures in a sequence of Bernoulli trials with probability of success p $(q=1-p)$, is $P(n;k)=C_{n+k-1}^{n-1}p^nq^k=C_{n+k-1}^{k}p^nq^k$.

Geometric Model: The probability of the 1^{st} success occurring after k failures in a sequence of Bernoulli trials with probability of success p $(q=1-p)$, is $p_k=pq^k$.

Sem. 4, Discrete Random Variables and Discrete Random Vectors

Bernoulli Distribution with parameter $p \in (0,1)$: $X \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$

Binomial Distribution with parameters $n \in \mathbb{N}, p \in (0,1)$: $X \begin{pmatrix} k \\ C_n^k p^k q^{n-k} \end{pmatrix}_{k=\overline{0,n}}$

Hypergeometric Distribution with parameters $N, n_1, n \in \mathbb{N}$, $\quad n, n_1 \le N$: $X \begin{pmatrix} k \\ p_k \end{pmatrix}_{k=\overline{0,n}}$, where $p_k = \dfrac{C_{n_1}^k C_{N-n_1}^{n-k}}{C_N^n}$

Poisson Distribution with parameter $\lambda > 0$: $X \begin{pmatrix} k \\ p_k \end{pmatrix}_{k \in \mathbb{N}}$, where $p_k = \dfrac{\lambda^k}{k!} e^{-\lambda}$

Pascal Distribution with parameters $n \in \mathbb{N}, p \in (0,1)$: $X \begin{pmatrix} k \\ C_{n+k-1}^k p^n q^k \end{pmatrix}_{k \in \mathbb{N}}$

Geometric Distribution with parameter $p \in (0,1)$: $X \begin{pmatrix} k \\ pq^k \end{pmatrix}_{k \in \mathbb{N}}$

Discrete Uniform Distribution with parameter $m \in \mathbb{N}$: $X \begin{pmatrix} k \\ \frac{1}{m} \end{pmatrix}_{k=\overline{1,m}}$

Cumulative Distribution Function $F_X : \mathbb{R} \to \mathbb{R}$, $F_X(x) = P(X < x)$

Discrete Random Vector: $(X, Y) : S \to \mathbb{R}^2$,
pdf $p_{ij} = P(X = x_i, Y = y_j)$, $(i,j) \in I \times J$,

cdf $F = F_{(X,Y)} : \mathbb{R}^2 \to \mathbb{R}$, $F(x,y) = P(X < x, Y < y) = \sum\limits_{x_i < x} \sum\limits_{y_j < y} p_{ij}$, $\forall (x,y) \in \mathbb{R}^2$,

$p_i = P(X = x_i) = \sum\limits_{j \in J} p_{ij}$, $\forall i \in I$, $q_j = P(Y = y_j) = \sum\limits_{i \in I} p_{ij}$, $\forall j \in J$ (marginal densities)

Operations: $X \begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i \in I}$, $Y \begin{pmatrix} y_j \\ q_j \end{pmatrix}_{j \in J}$
X and Y are **independent** <=> $p_{ij} = P(X = x_i, Y = y_j) = P(X = x_i) P(Y = y_j) = p_i q_j$.

$X + Y \begin{pmatrix} x_i + y_j \\ p_{ij} \end{pmatrix}_{(i,j) \in I \times J}$, $\alpha X \begin{pmatrix} \alpha x_i \\ p_i \end{pmatrix}_{i \in I}$, $XY \begin{pmatrix} x_i y_j \\ p_{ij} \end{pmatrix}_{(i,j) \in I \times J}$, $X/Y \begin{pmatrix} x_i / y_j \\ p_{ij} \end{pmatrix}_{(i,j) \in I \times J}$ $(y_j \neq 0)$

Sem. 5, Cont. R. Variables, Cont. R. Vectors, Functions of Cont. R. Variables

$X : S \to \mathbb{R}$ cont. random variable with pdf $f : \mathbb{R} \to \mathbb{R}$, cdf $F : \mathbb{R} \to \mathbb{R}$. Properties:

1. F is absolutely continuous and $F(x) = P(X < x) = \int\limits_{-\infty}^{x} f(t)dt$ **2.** $f(x) \ge 0$, $\forall x \in \mathbb{R}$, $\int\limits_{\mathbb{R}} f(x) = 1$

3. $P(X = x) = 0$, $\forall x \in \mathbb{R}$, $P(a < X < b) = \int\limits_a^b f(t)dt$ **4.** F is left continuous and increasing

5. $F(-\infty) = 0$, $F(\infty) = 1$

Continuous R. Vector: $(X, Y) : S \to \mathbb{R}^2$, pdf $f = f_{(X,Y)} : \mathbb{R}^2 \to \mathbb{R}$, cdf $F = F_{(X,Y)} : \mathbb{R}^2 \to \mathbb{R}$, $F(x,y) = P(X < x, Y < y) = \int\limits_{-\infty}^{x} \int\limits_{-\infty}^{y} f(u,v)\, dv\, du$, $\forall (x,y) \in \mathbb{R}^2$. Properties:

1. $P(a_1 \le X < b_1, a_2 \le Y < b_2) = F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2)$
2. F is left continuous and increasing in each variable **3.** $F(\infty, \infty) = 1$, $F(-\infty, y) = F(x, -\infty) = 0$, $\forall x, y \in \mathbb{R}$
4. X, Y independent <=> $F(x,y) = F_X(x) F_Y(y)$, <=> $f_{(X,Y)}(x,y) = f_X(x) f_Y(y)$, $\forall (x,y) \in \mathbb{R}^2$
5. $F_X(x) = F(x, \infty)$, $F_Y(y) = F(\infty, y)$, $\forall x, y \in \mathbb{R}$ (marginal cdf's) **6.** $P((X,Y) \in D) = \iint\limits_D f(x,y)\, dy\, dx$

7. $f_X(x) = \int\limits_{\mathbb{R}} f(x,y)dy,\ \forall x \in \mathbb{R},\ f_Y(y) = \int\limits_{\mathbb{R}} f(x,y)dx,\ \forall y \in \mathbb{R}$ (marginal densities)

8. Operations:

Sum: $f_{X+Y}(z) = \int\limits_{\mathbb{R}} f_{(X,Y)}(u, z-u)du \overset{X,Y ind}{=} \int\limits_{\mathbb{R}} f_X(u) f_Y(z-u)du$

Product: $f_{XY}(z) = \int\limits_{\mathbb{R}} f_{(X,Y)}\left(u, \frac{z}{u}\right)\frac{1}{|u|}du \overset{X,Y ind}{=} \int\limits_{\mathbb{R}} f_X(u) f_Y\left(\frac{z}{u}\right)\frac{1}{|u|}du$

Quotient: $f_{X/Y}(z) = \int\limits_{\mathbb{R}} f_{(X,Y)}(uz, u)\,|u|du \overset{X,Y ind}{=} \int\limits_{\mathbb{R}} f_X(uz) f_Y(u)\,|u|du$

Function $Y = g(X)$: $g : \mathbb{R} \to \mathbb{R}$ diff., $g' \neq 0$, strictly monotone $f_Y(y) = \dfrac{f_X\left(g^{-1}(y)\right)}{|g'\left(g^{-1}(y)\right)|}$, $y \in g(\mathbb{R})$

Sem. 6, Numerical Characteristics of Random Variables

Expectation:

X discr. with pdf $X\begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i\in I}$, $E(X) = \sum\limits_{i\in I} x_i p_i$, X cont. with pdf $f : \mathbb{R} \to \mathbb{R}$, $E(X) = \int\limits_{\mathbb{R}} x f(x)dx$.

Variance: $V(X) = E\left((X - E(X))^2\right) = E\left(X^2\right) - (E(X))^2$.

Standard Deviation: $\sigma(X) = \sqrt{V(X)}$.

Moments of order k:

- **initial** $\nu_k = E\left(X^k\right)$,
- **absolute** $\underline{\nu_k} = E\left(|X|^k\right)$,
- **central** $\mu_k = E\left((X - E(X))^k\right)$.

Covariance: $\mathrm{cov}(X,Y) = E\left((X - E(X))(Y - E(Y))\right) = E(XY) - E(X)E(Y)$

Correlation Coefficient: $\rho(X,Y) = \dfrac{\mathrm{cov}(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}}$

Properties:

1. $E(aX+b) = aE(X) + b$, $V(aX+b) = a^2V(X)$ **2.** $E(X+Y) = E(X) + E(Y)$

3. if X and Y are independent, then $E(XY) = E(X)E(Y)$ and $V(X+Y) = V(X) + V(Y)$

4. $h : \mathbb{R} \to \mathbb{R}$, X discrete, then $E(h(X)) = \sum\limits_{i\in I} h(x_i)p_i$, X continuous, then $E(h(X)) = \int\limits_{\mathbb{R}} h(x)f(x)dx$

5. $\mathrm{cov}(X,Y) = E(XY) - E(X)E(Y)$ **6.** $V\left(\sum\limits_{i=1}^n a_iX_i\right) = \sum\limits_{i=1}^n a_i^2 V(X_i) + 2\sum\limits_{1\le i<j\le n} a_ia_j\mathrm{cov}(X_i, X_j)$

7. X, Y independent $=>$ $\mathrm{cov}(X,Y) = \rho(X,Y) = 0$ (X and Y are *uncorrelated*)

8. $-1 \le \rho(X,Y) \le 1$; $\rho(X,Y) = \pm 1 \iff \exists\, a,b \in \mathbb{R},\ a \neq 0$ s.t. $Y = aX + b$

9. (X,Y) a cont. r. vector with pdf $f(x,y)$, $h : \mathbb{R}^2 \to \mathbb{R}^2$, then $E(h(X,Y)) = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} h(x,y)f(x,y)dxdy$.

Sem. 7, Inequalities, Sequences of Random Variables

Hölder's Inequality: $E(|XY|) \le \left(E(|X|^p)\right)^{\frac{1}{p}} \cdot \left(E(|Y|^q)\right)^{\frac{1}{q}}$, $\forall p, q > 1$, $\frac{1}{p} + \frac{1}{q} = 1$.

Markov's Inequality: $P(|X| \ge a) \le \frac{1}{a}E(|X|)$, $\forall a > 0$.

Chebyshev's Inequality: $P(|X - E(X)| \ge \epsilon) \le \dfrac{V(X)}{\epsilon^2}$, $\forall \epsilon > 0$.

Convergence:

1) **in probability** $X_n \xrightarrow{p} X$, if $\lim\limits_{n\to\infty} P\Big(|X_n - X| < \varepsilon\Big) = 1, \quad \forall\varepsilon > 0$;

2) **strongly** $X_n \xrightarrow{s} X$, if $\lim\limits_{n\to\infty} P\Big(\bigcap\limits_{k\ge n}\{|X_k - X| < \varepsilon\}\Big) = 1, \quad \forall\varepsilon > 0$;

3) **almost surely** $X_n \xrightarrow{a.s.} X$, if $P\Big(\lim\limits_{n\to\infty} X_n = X\Big) = 1$;

4) **in distribution** $X_n \xrightarrow{d} X$, if $\lim\limits_{n\to\infty} F_n(x) = F(x)$, $\forall x \in \mathbb{R}$ continuity point for F;

5) **in mean of order** r, $0 < r < \infty$ $X_n \xrightarrow{L^r} X$, if $\lim\limits_{n\to\infty} E(|X_n - X|^r) = 0$.

Properties 1. 2) $<=>$ 3) $=>$ 1) $=>$ 4) **2.** 5) $=>$ 1).

STATISTICS

X a population characteristic, $X_1, X_2, ..., X_n$ a sample of size n, i.e. independent and identically distributed, with the same pdf as X; θ target parameter, $\overline{\theta} = \overline{\theta}(X_1, X_2, ..., X_n)$ point estimator.

Sample Mean: $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$,

Sample Moment: $\overline{\nu_k} = \frac{1}{n}\sum_{i=1}^{n} X_i^k$,

Sample Absolute Moment: $\overline{\mu_k} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k$,

Sample Variance: $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$.

Likelihood Function of a Sample: $L(X_1, ..., X_n|\theta) = \prod_{i=1}^{n} f(X_i|\theta)$.

Fisher's Information: $I_n(\theta) = E\left[\left(\frac{\partial \ln L(X_1, ..., X_n|\theta)}{\partial\theta}\right)^2\right]$.

- if the range of X does not depend on θ, then $I_n(\theta) = -E\left[\frac{\partial^2 \ln L(X_1, ..., X_n|\theta)}{\partial^2\theta}\right]$ and $I_n(\theta) = nI_1(\theta)$.

Efficiency of an Absolutely Correct Estimator: $e(\overline{\theta}) = \frac{1}{I_n(\theta)V(\overline{\theta})}$.

Estimator $\overline{\theta}$ is
- **unbiased**: $E(\overline{\theta}) = \theta$;
- **MVUE**: $E(\overline{\theta}) = \theta$ and $V(\overline{\theta}) \leq V(\hat{\theta})$, $\forall\, \hat{\theta}$ unbiased estimator;
- **absolutely correct**: $E(\overline{\theta}) = \theta$ and $\lim_{n\to\infty} V(\overline{\theta}) = 0$;
- **efficient**: absolutely correct and $e(\overline{\theta}) = 1$.

Statistic $S = S(X_1, X_2, ..., X_n)$ is
- **sufficient for** θ: the cond. pdf $f(X_1, ..., X_n|S)$ does not depend on θ $\overset{\text{Fact.Crit.}}{<=>}$ $L(x_1, ..., x_n|\theta) = g(s, \theta)h(x_1, ..., x_n)$;
- **complete for the family of distributions** $f(x \mid \theta),\ \theta \in A$: $\quad E(\varphi(S)) = 0, \forall\, \theta \in A \ => \ \varphi \overset{\text{a.s.}}{=} 0$.

Method of Moments:
Solve the system $\nu_k = \overline{\nu_k}$ for all unknown parameters.

Method of Maximum Likelihood:

Solve the system $\frac{\partial L(X_1, ..., X_n|\theta)}{\partial\theta_j} = 0$ or $\frac{\partial \ln L(X_1, ..., X_n|\theta)}{\partial\theta_j} = 0,\ j = \overline{1, m}$ for the unknown parameters $\theta = (\theta_1, ..., \theta_m)$.

Lehmann-Scheffé Theorem: Let $\hat{\theta}$ be an unbiased estimator and S a sufficient and complete statistic for θ. Then $\overline{\theta} = E(\hat{\theta}|S)$ is an MVUE.

Rao-Cramer Inequality: Let $\overline{\theta}$ be an absolutely correct estimator for θ. Then $V(\overline{\theta}) \geq \frac{1}{I_n(\theta)}$.

Hypothesis Testing: $H_0 : \theta = \theta_0$ with one of the alternatives $H_1 : \begin{cases} \theta < \theta_0 & \text{(left-tailed test)}, \\ \theta > \theta_0 & \text{(right-tailed test)}, \\ \theta \neq \theta_0 & \text{(two-tailed test)}. \end{cases}$

Significance Level: $\alpha = P(\text{ type I error}) = P(\text{ reject } H_0 \mid H_0) = P(TS \in RR \mid \theta = \theta_0)$.
Type II Error: $\beta = P(\text{ type II error}) = P(\text{ accept } H_0 \mid H_1) = P(TS \notin RR\ |H_1)$.
Power of a Test: $\pi(\theta^*) = P(\text{ reject } H_0 \mid \theta = \theta^*) = P(TS \in RR \mid \theta = \theta^*)$.
Neyman-Pearson Lemma (NPL): Suppose we test two simple hypotheses $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$. Let $L(\theta^*)$ denote the likelihood function of the sample, when $\theta = \theta^*$. Then for every $\alpha \in (0, 1)$, a most powerful test (a test that maximizes the power $\pi(\theta_1)$) is the test with $RR = \left\{\frac{L(\theta_1)}{L(\theta_0)} \geq k_\alpha\right\}$, for some constant $k_\alpha > 0$.