PROBABILITY THEORY

Sem. 1, Euler's Functions; Counting, Outcomes, Events

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.** $\beta(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$, $\forall a,b > 0$.

Arrangements: $A_n^k = \frac{n!}{(n-k)!}$; **Permutations:** $P_n = A_n^n = n!$; **Combinations:** $C_n^k = \frac{n!}{k!(n-k)!}$.

Sem. 2, Class. Probability; Rules of Probability; Cond. Probability; Ind. Events

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

Mutually Exclusive Events: A, B **m. e. (disjoint, incompatible)** $<=>$ $P(A\cap B) = 0$.

Rules of Probability:
$P(\overline{A}) = 1 - P(A)$;
$P(A\cup B) = P(A) + P(B) - P(A\cap B)$;
$P(A\setminus B) = P(A) - P(A\cap B)$.

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

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

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

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

Sem. 3, Probabilistic Models

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

<u>**Hypergeometric Model**</u>: 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 $P(n;k) = \frac{C_{n_1}^k C_{N-n_1}^{n-k}}{C_N^n}$

<u>**Poisson Model**</u>: 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 $P(n;k) = \sum\limits_{1\le i_1<\dots<i_k\le n} p_{i_1}\dots p_{i_k}q_{i_{k+1}}\dots q_{i_n},\quad i_{k+1},\dots,i_n \in \{1,\dots,n\}\setminus\{i_1,\dots,i_k\}$ = the coefficient of x^k in the expansion $(p_1x+q_1)(p_2x+q_2)\dots(p_nx+q_n)$.

<u>**Pascal (Negative Binomial) Model**</u>: 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 $P(n;k) = C_{n+k-1}^{n-1}p^nq^k = C_{n+k-1}^k p^nq^k$.

<u>**Geometric Model**</u>: 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 $n \in \mathbb{N}, p \in (0,1)$: $X \begin{pmatrix} k \\ C_n^k p^k q^{n-k} \end{pmatrix}_{k=\overline{0,n}}$

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

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

X represents the number of "rare events" that occur in a fixed period of time; λ represents the frequency, the average number of events occurring per time unit.

(Negative Binomial) 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 $p \in (0,1)$: $X \begin{pmatrix} k \\ pq^k \end{pmatrix}_{k \in \mathbb{N}}$

Cumulative Distribution Function (cdf) $F_X : \mathbb{R} \to \mathbb{R}$, $F_X(x) = P(X \le x) = \sum\limits_{x_i \le x} p_i$

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

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

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

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** $<=>$ $p_{ij} = P(X = x_i, Y = y_j) = P(X = x_i)\, P(Y = y_j) = p_i q_j$.

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

Sem. 5, Continuous Random Variables and Continuous Random Vectors

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

1. $F(x) = P(X \le x) = \int\limits_{-\infty}^{x} f(t)dt$
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. $F(-\infty) = 0$, $F(\infty) = 1$

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

1. $P(a_1 < X \le b_1, a_2 < Y \le b_2) = F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2)$

2. $F(\infty,\infty)=1,\ F(-\infty,y)=F(x,-\infty)=0,\ \forall x,y\in\mathbb{R}$
3. $F_X(x)=F(x,\infty),\ F_Y(y)=F(\infty,y),\ \forall x,y\in\mathbb{R}$ (marginal cdf's)
4. $P((X,Y)\in D)=\iint\limits_D f(x,y)\ dy\ dx$
5. $f_X(x)=\int\limits_{\mathbb{R}} f(x,y)dy,\ \forall x\in\mathbb{R},\ f_Y(y)=\int\limits_{\mathbb{R}} f(x,y)dx,\ \forall y\in\mathbb{R}$ (marginal densities)
6. X and Y are independent $<=> f_{(X,Y)}(x,y)=f_X(x)f_Y(y),\ \forall(x,y)\in\mathbb{R}^2$.

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

Uniform distribution $\mathcal{U}(a,b),\ -\infty<a<b<\infty$: pdf $f(x)=\dfrac{1}{b-a}, x\in[a,b]$.

Normal distribution $N(\mu,\sigma),\mu\in\mathbb{R},\sigma>0$: pdf $f(x)=\dfrac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x\in\mathbb{R}$.

Gamma distribution $Gamma(a,b),\ a,b>0$: pdf $f(x)=\dfrac{1}{\Gamma(a)b^a}x^{a-1}e^{-\frac{x}{b}}\ ,\ x>0$.

Exponential distribution $Exp(\lambda)=Gamma(1,1/\lambda),\ \lambda>0$: pdf $f(x)=\lambda e^{-\lambda x}, x>0$.
- Exponential distribution models *time*: waiting time, interarrival time, failure time, time between rare events, etc. The parameter λ represents the frequency of rare events, measured in time^{-1}.
- Gamma distribution models the *total* time of a multistage scheme.
- For $\alpha\in\mathbb{N}$, a $Gamma(\alpha,1/\lambda)$ variable is the sum of α independent $Exp(\lambda)$ variables.

Sem. 6, Numerical Characteristics of Random Variables

Expectation:
X discr. with pdf $X\begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i\in I}$, $E(X)=\sum\limits_{i\in I}x_ip_i$, X cont. with pdf $f:\mathbb{R}\to\mathbb{R}$, $E(X)=\int\limits_{\mathbb{R}}xf(x)dx$.
Variance: $V(X)=E\left((X-E(X))^2\right)=E\left(X^2\right)-(E(X))^2$.
Standard Deviation: $\sigma(X)=\sqrt{V(X)}$.
Moment of order k $\nu_k=E\left(X^k\right)$,
Absolute moment of order k $\underline{\nu_k}=E\left(|X|^k\right)$,
Central moment of order k $\mu_k=E\left((X-E(X))^k\right)$.
Covariance: $\operatorname{cov}(X,Y)=E\left((X-E(X))(Y-E(Y))\right)=E(XY)-E(X)E(Y)$
Correlation Coefficient: $\rho(X,Y)=\dfrac{\operatorname{cov}(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}}$

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 $E\left(h(X)\right)=\sum\limits_{i\in I}h(x_i)p_i$, X continuous, then $E\left(h(X)\right)=\int\limits_{\mathbb{R}}h(x)f(x)dx$
5. $\operatorname{cov}(X,Y)=E(XY)-E(X)E(Y)$
6. $V\left(\sum\limits_{i=1}^n a_iX_i\right)=\sum\limits_{i=1}^n a_i^2V(X_i)+2\sum\limits_{1\leq i<j\leq n}a_ia_j\operatorname{cov}(X_i,X_j)$
7. X,Y independent $=> \operatorname{cov}(X,Y)=\rho(X,Y)=0$ (X and Y are *uncorrelated*)
8. $-1\leq\rho(X,Y)\leq 1;\ \rho(X,Y)=\pm1\ <=>\ \exists\ a,b\in\mathbb{R},\ a\neq 0$ 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 $E\left(h(X,Y)\right)=\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}h(x,y)f(x,y)dxdy$.

Sem. 7, Inequalities; Central Limit Theorem; Markov Chains; Point Estimators

Markov's Inequality: $P\left(|X|\geq a\right)\leq\dfrac{1}{a}E\left(|X|\right), \forall a>0$.

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

<u>Central Limit Theorem</u>(CLT) Let $X_1, \ldots, X_n$ be independent random variables with the same expectation $\mu = E(X_i)$ and same standard deviation $\sigma = \sigma(X_i)$ and let $S_n = \sum_{i=1}^{n} X_i$. Then, as $n \to \infty$,

$$Z_n = \frac{S_n - E(S_n)}{\sigma(S_n)} = \frac{S_n - n\mu}{\sigma\sqrt{n}} \longrightarrow Z \in N(0,1), \text{ in distribution (in cdf).}$$

<u>STATISTICS</u>

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: $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$,

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

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

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

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

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

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

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

<u>Estimator</u> $\overline{\theta}$ is

- **unbiased**: $E(\overline{\theta}) = \theta$;
- **MVUE (min. var. unbiased estimator)**: $E(\overline{\theta}) = \theta$ 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 $e(\overline{\theta}) = 1$.

<u>Method of Moments</u>:

Solve the system $\nu_k = \overline{\nu}_k$, for as many parameters as needed ($k = 1 \ldots$ nr. of unknown parameters).

<u>Method of Maximum Likelihood</u>:

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

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

<u>Significance Level</u>: $\alpha = P(\text{type I error}) = P(\text{reject } H_0 \mid H_0) = P(TS \in RR \mid \theta = \theta_0)$.

<u>Type II Error</u>: $\beta = P(\text{type II error}) = P(\text{do not reject } H_0 \mid H_1) = P(TS \notin RR \mid H_1)$.

<u>Power of a Test</u>: $\pi(\theta^*) = P(\text{reject } H_0 \mid \theta = \theta^*) = P(TS \in RR \mid \theta = \theta^*)$.

<u>Neyman-Pearson Lemma (NPL)</u>: 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 $RR = \left\{\dfrac{L(\theta_1)}{L(\theta_0)} \geq k_\alpha\right\}$, for some constant $k_\alpha > 0$.