

PROBABILITY THEORY

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

Euler's Gamma Function $\Gamma : (0, \infty) \rightarrow (0, \infty)$ $\Gamma(a) = \int_0^{\infty} x^{a-1} e^{-x} dx$

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. $\Gamma\left(\frac{1}{2}\right) = \sqrt{2} \int_0^{\infty} e^{-\frac{t^2}{2}} dt = \int_{\mathbb{R}} e^{-t^2} dt = \sqrt{\pi}$.

Euler's Beta Function $\beta : (0, \infty) \times (0, \infty) \rightarrow (0, \infty)$ $\beta(a, b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx$

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) $\Leftrightarrow P(A \cap B) = 0$.

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. $\Leftrightarrow P(A \cap B) = P(A)P(B) \Leftrightarrow P(A|B) = P(A)$.

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

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

Sem. 3, Probabilistic Models

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

Hypergeometric Model: The probability that in n trials, we get k white balls out of n_1 and $n - k$ black balls out of $N - n_1$ ($0 \leq k \leq n_1$, $0 \leq n - k \leq N - n_1$), is $P(n; k) = \frac{C_{n_1}^k C_{N-n_1}^{n-k}}{C_N^n}$

Poisson Model: The probability of k successes ($0 \leq k \leq n$) in n trials, with probability of success p_i in the i^{th} trial ($q_i = 1 - p_i$), $i = \overline{1, n}$, is $P(n; k) = \sum_{1 \leq i_1 < \dots < i_k \leq n} p_{i_1} \dots p_{i_k} q_{i_{k+1}} \dots q_{i_n}$, $i_{k+1}, \dots, i_n \in \{1, \dots, n\} \setminus \{i_1, \dots, i_k\}$ = the coefficient of x^k in the expansion $(p_1 x + q_1)(p_2 x + q_2) \dots (p_n x + q_n)$.

Pascal (Negative Binomial) 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 $P(n; k) = C_{n+k-1}^{n-1} p^n q^k = C_{n+k-1}^k p^n q^k$.

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 $p_k = p q^k$.

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

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

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

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

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) $F_X : \mathbb{R} \rightarrow \mathbb{R}, F_X(x) = P(X \leq x) = \sum_{x_i \leq x} p_i$

Discrete Random Vector: $(X, Y) : S \rightarrow \mathbb{R}^2$,

– **(joint) pdf** $p_{ij} = P(X = x_i, Y = y_j), (i, j) \in I \times J$,

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

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

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

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

Sem. 5, Continuous Random Variables and Continuous Random Vectors

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

1. $F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt$

2. $f(x) \geq 0, \forall x \in \mathbb{R}, \int_{\mathbb{R}} f(x) = 1$

3. $P(X = x) = 0, \forall x \in \mathbb{R}, P(a < X < b) = \int_a^b f(t) dt$

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

Continuous R. Vector: $(X, Y) : S \rightarrow \mathbb{R}^2$, pdf $f = f_{(X,Y)} : \mathbb{R}^2 \rightarrow \mathbb{R}$, cdf $F = F_{(X,Y)} : \mathbb{R}^2 \rightarrow$

$\mathbb{R}, F(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) dv du, \forall (x, y) \in \mathbb{R}^2$. Properties:

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

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. $P((X, Y) \in D) = \int_D \int f(x, y) dy dx$
5. $f_X(x) = \int_{\mathbb{R}} f(x, y) dy, \forall x \in \mathbb{R}, f_Y(y) = \int_{\mathbb{R}} f(x, y) dx, \forall y \in \mathbb{R}$ (marginal densities)
6. X and Y are independent $\Leftrightarrow f_{(X,Y)}(x, y) = f_X(x)f_Y(y), \forall (x, y) \in \mathbb{R}^2$.

Function $Y = g(X)$: X r.v., $g: \mathbb{R} \rightarrow \mathbb{R}$ differentiable with $g' \neq 0$, strictly monotone

$$f_Y(y) = \frac{f_X(g^{-1}(y))}{|g'(g^{-1}(y))|}, y \in g(\mathbb{R})$$

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

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

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

Exponential distribution $\text{Exp}(\lambda) = \text{Gamma}(1, 1/\lambda), \lambda > 0$: 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 $\text{Gamma}(\alpha, 1/\lambda)$ variable is the sum of α independent $\text{Exp}(\lambda)$ variables.

Sem. 6, Numerical Characteristics of Random Variables

Expectation:

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

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

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

Moment of order k $\nu_k = E(X^k)$,

Absolute moment of order k $\underline{\nu}_k = E(|X|^k)$,

Central moment of order k $\mu_k = E((X - E(X))^k)$.

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

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

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. $h: \mathbb{R} \rightarrow \mathbb{R}, X$ discrete, then $E(h(X)) = \sum_{i \in I} h(x_i) p_i$, X continuous, then $E(h(X)) = \int_{\mathbb{R}} h(x) f(x) dx$

5. $\text{cov}(X, Y) = E(XY) - E(X)E(Y)$

6. $V\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 V(X_i) + 2 \sum_{1 \leq i < j \leq n} a_i a_j \text{cov}(X_i, X_j)$

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

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

9. (X, Y) a cont. r. vector with pdf $f(x, y), h: \mathbb{R}^2 \rightarrow \mathbb{R}^2$, then $E(h(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) f(x, y) dx dy$.

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

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

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

Central Limit Theorem (CLT) Let X_1, \dots, 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 $n \rightarrow \infty$,

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

STATISTICS

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

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

Sample Moment: $\bar{\nu}_k = \frac{1}{n} \sum_{i=1}^n X_i^k,$

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

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

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

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

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

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

Estimator $\bar{\theta}$ is

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

Method of Moments:

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

Method of Maximum Likelihood:

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

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: $\alpha = P(\text{type I error}) = P(\text{reject } H_0 | H_0) = P(TS \in RR | \theta = \theta_0).$

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

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

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 $k_\alpha > 0$.