Theory Review

Seminar 6

Expectation:

- if $X \begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i \in I}$ is discrete, then $E(X) = \sum_{i \in I} x_i p_i$.

- if X is continuous with pdf f, then $E(X) = \int_{\mathbb{R}} x f(x) dx$.

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

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

Moments:

- **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**: $\mu_k = E\left((X - E(X))^k\right)$.

Properties:

1. $E(aX + b) = aE(X) + b$, $V(aX + b) = a^2 V(X)$
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. if $h : \mathbb{R} \to \mathbb{R}$ is a measurable function, X a random variable;

- if X is discrete, then $E(h(X)) = \sum_{i \in I} h(x_i) p_i$

- if X is continuous, then $E(h(X)) = \int_{\mathbb{R}} h(x) f(x) dx$

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

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

Properties:

1. $\mathrm{cov}(X, Y) = E(XY) - E(X)E(Y)$
2. $V\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 V(X_i) + 2 \sum_{1 \le i < j \le n} a_i a_j \mathrm{cov}(X_i, X_j)$
3. X, Y independent $=>$ $\mathrm{cov}(X, Y) = \rho(X, Y) = 0$ (X and Y are *uncorrelated*)
4. $-1 \le \rho(X, Y) \le 1$; $\rho(X, Y) = \pm 1 \iff \exists\, a, b \in \mathbb{R},\ a \neq 0$ s.t. $Y = aX + b$

Let (X, Y) be a continuous random vector with pdf $f(x, y)$, let $h : \mathbb{R}^2 \to \mathbb{R}^2$ a mea

$$E\left(h(X, Y)\right) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) f(x, y) dx dy$$

Theory Review

Seminar 7

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

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

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 $\sigma = \sigma(X_i) = \mathrm{Std}(X_i)$ and let $S_n = \sum_{i=1}^{n} X_i$. Then, as $n \to \infty$,

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

Point Estimators

- method of moments: solve the system $\nu_k = \overline{\nu}_k$, for as many parameters as needed ($k = 1, \dots,$ nr. of unknown parameters);
- method of maximum likelihood: solve $\dfrac{\partial \ln L(X_1, \dots, X_n; \theta)}{\partial \theta_j} = 0$, where $L(X_1, \dots, X_n; \theta) = \prod_{i=1}^{n} f(X_i; \theta)$ is the likelihood function;
- **standard error** of an estimator $\overline{\theta}$: $\sigma_{\hat{\theta}} = \sigma(\overline{\theta}) = \sqrt{V(\overline{\theta})}$;
- **Fisher information** $I_n(\theta) = -E\left[\dfrac{\partial^2 \ln L(X_1, \dots, X_n; \theta)}{\partial \theta^2}\right]$; if the range of X does not depend on θ, then $I_n(\theta) = n I_1(\theta)$;
- **efficiency** of an absolutely correct estimator $\overline{\theta}$: $e(\overline{\theta}) = \dfrac{1}{I_n(\theta) V(\overline{\theta})}$.
- an estimator $\overline{\theta}$ for the target parameter θ is
  - **unbiased**, if $E(\overline{\theta}) = \theta$;
  - **absolutely correct**, if $E(\overline{\theta}) = \theta$ and $V(\overline{\theta}) \to 0$, as $n \to \infty$;
  - **MVUE** (minimum variance unbiased estimator), if $E(\overline{\theta}) = \theta$ and $V(\overline{\theta}) \le V(\hat{\theta}),\ \forall \hat{\theta}$ unbiased estimator;
  - **efficient**, if $e(\overline{\theta}) = 1$.

- $\overline{\theta}$ efficient $=>$ $\overline{\theta}$ MVUE.

Lemma 6.4 (Neyman-Pearson (NPL)). *Let X be a characteristic with pdf $f(x; \theta)$, with $\theta \in A \subset \mathbb{R}$, unknown. Suppose we test on θ the simple hypotheses*

$$H_0 : \quad \theta = \theta_0$$
$$H_1 : \quad \theta = \theta_1,$$

based on a random sample $X_1, \dots, X_n$. Let $L(\theta) = L(X_1, \dots, X_n; \theta)$ denote the likelihood function of this sample. Then for a fixed $\alpha \in (0, 1)$, a most powerful test is the test with rejection region given by

$$RR = \left\{ \frac{L(\theta_1)}{L(\theta_0)} \ge k_\alpha \right\}, \qquad (6.5)$$

where the constant $k_\alpha > 0$ depends only on α and the sample variables.

$\beta(\mu_1) = P(\text{not reject } H_0 \mid H_1)$

$$\pi(\theta_1) = 1 - \beta(\theta_1)$$

Theory Review

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

Seminar 1

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: $\beta : (0,\infty)\times(0,\infty) \to (0,\infty), \beta(a,b) = \int\limits_0^1 x^{a-1}(1-x)^{b-1}dx.$

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

Theory Review

Seminar 3

<u>Binomial Model</u>: The probability of k successes in n Bernoulli trials, with probability of success p $(q = 1-p)$, 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 successes out of n_1 and $n-k$ failures 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

$$P(n;k) = \sum_{1\le i_1<\ldots<i_k\le n} p_{i_1}\ldots p_{i_k}q_{i_{k+1}}\ldots q_{i_n},\qquad i_{k+1},\ldots,i_n \in \{1,\ldots,n\}\setminus\{i_1,\ldots,i_k\}$$
$$= \text{the coefficient of } x^k \text{ in the polynomial expansion } (p_1x+q_1)(p_2x+q_2)\ldots(p_nx+q_n).$$

<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

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

<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

$p_k = pq^k.$

<u>(Negative Binomial) Pascal Distribution</u> with parameters $n \in \mathbb{N}, p \in (0,1)$ pdf:

$$X\begin{pmatrix} k \\ C_{n+k-1}^k p^n q^k \end{pmatrix}_{k=0,1,\ldots}$$

<u>Geometric Distribution</u> with parameter $p \in (0,1)$ pdf: $X\begin{pmatrix} k \\ pq^k \end{pmatrix}_{k=0,1,\ldots}$

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

$(X,Y) : S \to \mathbb{R}^2$ **discrete random vector**:

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

For $X\begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i\in I}, Y\begin{pmatrix} y_j \\ q_j \end{pmatrix}_{j\in J},$

X and Y are **independent** $<=> p_{ij} = P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j) = p_iq_j.$

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

Theory Review

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

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

Seminar 2

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

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

Theory Review

Seminar 4

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

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

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

<u>Hypergeometric Distribution</u> with parameters $N, n_1, n \in \mathbb{N}$ $(n_1 \le N)$ pdf: $X\begin{pmatrix} k \\ \frac{C_{n_1}^k C_{N-n_1}^{n-k}}{C_N^n} \end{pmatrix}_{k=\overline{0,n}}$

<u>Poisson Distribution</u> with parameter $\lambda > 0$ pdf: $X\begin{pmatrix} k \\ \frac{\lambda^k}{k!}e^{-\lambda} \end{pmatrix}_{k=0,1,\ldots}$

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

Theory Review

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

Seminar 5

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. $P(X = x) = 0, \forall x \in \mathbb{R},\ P(a < X < b) = P(a \le X \le b) = \int\limits_a^b f(t)dt$
4. $F(-\infty) = 0,\ F(\infty) = 1$

$(X,Y) : S \to \mathbb{R}^2$ continuous random vector with pdf $f = f_{(X,Y)} : \mathbb{R}^2 \to \mathbb{R}$ and

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

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. $P((X,Y) \in D) = \int\int\limits_D f(x,y)\,dy\,dx$
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 $<=> f_{(X,Y)}(x,y) = f_X(x)f_Y(y),\ \forall (x,y) \in \mathbb{R}^2.$