Repetition of basic statistics

MVE220

Helga Kristín Ólafsdóttir

Mars 27, 2023

Problems are translated from [1]

Independent Events

Assume $P(B) > 0$. Then the **conditional probability** of A given B is

$$P(A|B) := \frac{P(A \cap B)}{P(B)}$$

A and B are **independent** if

$$P(A \cap B) = P(A)P(B)$$

Problem 1 - Independent Events (2.5.5. in [1])

Two dice are thrown. Assume their sum was 4. What is the conditional probability that

a) The first dice showed 3
b) The second dice showed 2 or less
c) Both dice showed an odd number

Distribution function

A random variable X can be discrete or continuous.

- If X discrete: **probability function** $p_X(x) := P(X = x)$, $x = x_1, x_2, ...$
- If X continuous: **probability density function (pdf)** f_X s.t. $P(X \in A) = \int_A f_X(t)dt$

In both cases: **(cumulative) distribution function (cdf)**

$$F_X(t) := P(X \leq t), \;\; -\infty < t < \infty$$

For the cdf we have

- $F(t) \in [0, 1]$
- $F(t) \underset{t \to -\infty}{\longrightarrow} 0$, and $F(t) \underset{t \to \infty}{\longrightarrow} 1$
- $P(a < X \leq b) = F(b) - F(a)$
- $P(X > a) = 1 - F(a)$

Problem 2 - Distribution function (3.3.1. in [1])

Let Y be a random variable with the cdf

$$F(t) = \begin{cases} 0 & t < 0 \\ t^2 & 0 \leq t \leq 1 \\ 1 & t > 1 \end{cases}$$

a) Sketch $F(t)$
b) Calculate $P(Y \leq 0.5)$
c) Calculate $P(0.5 < Y \leq 0.9)$

Expected value etc.

Expected value

$E[X] = \sum_k kP(X = k)$ if X discrete

$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$ if X continuous

$E[g(X)]$

$E[g(X)] = \sum_k g(k)P(X = k)$ if X discrete

$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$ if X continuous

k:th Moment

$E[X^k]$

Expected value etc.

Variance

$\sigma^2 = Var(X) := E[(X - E[X])^2] = E[X^2] - E[X]^2$

Standard deviation

$\sigma = \sqrt{Var(X)}$

Problem 3 - Expected value etc. (3.5.1. in [1])

Calculate the expectation and standard deviation of X.

k	1	2	3	4	5
$p_X(k)$	0.2	0.1	0.3	0.1	0.3

Correlation

Covariance

$Cov(X, Y) := E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$

Correlation coefficient

$\rho(X, Y) := \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$

X, Y are **uncorrelated** if $Cov(X, Y) = 0$

Problem 4 - Correlation (3.8.3. in [1])

Let (X, Y) be given with pdf $f_{X,Y}(x, y) = x + y$, $0 \leq x \leq 1$, $0 \leq y \leq 1$. Calculate $E[X]$, $E[Y]$, $Var(X)$, $Var(Y)$, $Cov(X, Y)$ and $\rho(X, Y)$.

Poisson processes

Poisson distribution

$X \sim Po(\lambda)$ $P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, ...$

$E[X] = \lambda, \quad Var(X) = \lambda$

Stochastic process

$\{X(t), t \in I\}$ family of random variables with index t from the index set I.

Poisson process

Let $\{U_k, k = 1, 2, 3...\}$ be a sequence of i.i.d. $Exp(\lambda)$ distributed random variables.

$T_n := \sum_{k=1}^{n} U_k$ is the time of the n-th event.

Define $N(t) :=$ events in $(0, t], \quad N(0) = 0$.

Then $\{N(t), t \geq 0\}$ is a **Poisson process** with intensity λ.

Poisson processes

Independent increments of Poisson process

$N(s+t) - N(s) \sim Po(\lambda t)$

$N(t) \sim Po(\lambda t)$

Problem 5 - Poisson processes (4.3.1. in [1])

Let $\{N(t), t \geq 0\}$ be a Poisson process with $\lambda = 2$. Calculate

a) $P(N(1) = 0)$
b) $P(N(3) = 4)$
c) $P(N(2) \leq 3)$
d) $P(N(0.5) > 1)$

Point estimation

Let $x_1, x_2, ..., x_n$ be observations of X that follows a distribution with some unknown parameter θ. Then we can **estimate** θ with θ^*

Maximum likelihood method

We want to find θ^* that maximises the **Likelihood function**

$$L(\theta) = \begin{cases} \prod_{i=1}^{n} p(x_i;\theta) & X \text{ discrete} \\ \prod_{i=1}^{n} f(x_i;\theta) & X \text{ continuous} \end{cases}$$

Easier to handle the **log-likelihood function**

$$l(\theta) = ln(L(\theta)) = \begin{cases} \sum_{i=1}^{n} ln(p(x_i;\theta)) & X \text{ discrete} \\ \sum_{i=1}^{n} ln(f(x_i;\theta)) & X \text{ continuous} \end{cases}$$

Problem 6 - Point estimation (7.2.9. in [1])

Nine observations were obtained from a distribution with pdf

$$f(x) = \frac{x}{\theta^2} e^{x/\theta}$$

Find the ML-estimate of θ.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
2.2	5.5	1.7	1.3	3.5	3.2	0.6	3.8	1.9

Confidence intervals

Given observations $\mathbf{x}$ we want to compute a $(1-\alpha)\%$ **confidence interval** $I_\theta = (\theta_1(\mathbf{x}), \theta_2(\mathbf{x}))$, where $P(\theta_1(\mathbf{X}) < \theta < \theta_2(\mathbf{X})) = 1-\alpha$

Problem 7 - Confidence intervals (7.3.1. in [1])

The random variable X is Poisson distributed with expected value μ. The 95% confidence interval for μ is $I_\mu = (0.8, 2.0)$. Calculate the 95% confidence interval for $p := P(X = 0)$.

References

[1] Sven Erick Alm and Tom Britton. *Stokastik: sannolikhetsteori och statistikteori med tillämpningar*. Liber, 2008.