TECHNISCHE UNIVERSITEIT EINDHOVEN
Faculteit Wiskunde en Informatica

Exam Pre-master Probability and Statistics (2DI50) on 11 May 2007, 9.00-12.00

- At the exam it is allowed to use the 'Statistisch Compendium' and a pocket calculator.
- The exam has 8 exercises with 20 items.
- The answers have to be in ENGLISH. Give not just answers, but also arguments.
- For each item one get get two points. In total one can get 40 points.

Problem 1:

Two balls are extracted from an urn containing white and black balls. After the first extraction, the selected ball is replaced in the urn. The probability of extracting a white ball is thus constant at each trial and it will be denoted by p, with $0 < p < 1$. Define the following two discrete random variables:

$$X = \begin{cases} 1 & \text{if the two extracted balls have the same color} \\ 0 & \text{otherwise} \end{cases}$$

$$Y = \text{(number of extracted white balls) - (number of extracted black balls)}$$

(a) Compute the joint probability mass function $f_{X,Y}(x, y) = \mathbb{P}(X = x, Y = y)$

$P(X = x, Y = y)$			Y		$P(X = x)$
		.	.	.	
X	.	.	.	.	.
	.	.	.	.	.
$P(Y = y)$		.	.	.	.

(b) Find the values of p for which X and Y are uncorrelated, namely the covariance is zero.

(c) Decide whether X and Y independent for the values of p found in the previous item.

Solution:
(a) Denote by $(w,w),(w,b),(b,w),(b,b)$ the results of the two extractions, where w stands for a white ball and b stands for a black ball. Then we have

$$X = \begin{cases} 1 & \text{for } (w,w) \text{ or } (b,b) \\ 0 & \text{for } (w,b) \text{ or } (w,b) \end{cases}$$

and

$$Y = \begin{cases} -2 & \text{for } (b,b) \\ 0 & \text{for } (w,b) \text{ or } (w,b) \\ 2 & \text{for } (w,w) \end{cases}$$

From this we see that $f_{X,Y}(1,-2) = (1-p)^2$, $f_{X,Y}(0,0) = 2p(1-p)$, $f_{X,Y}(1,2) = p^2$, and all the others $f_{X,Y}(x,y) = 0$.
(b) We look for $0 < p < 1$ such that

$$\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0\,.$$

We have

$$\mathbb{E}(XY) = -2(1-p)^2 + 2p^2 = 2(2p-1)$$
$$\mathbb{E}(X) = p^2 + (1-p)^2$$
$$\mathbb{E}(Y) = -2(1-p)^2 + 2p^2 = 2(2p-1)$$

The only possible solution is $p = 1/2$.
(c) X and Y are not independent. To conclude this it is enough to observe that $f_{X,Y}(0,-2) = 0$, $f_X(0) = 1/2$, $f_Y(-2) = -1/4$. This implies that $f_{X,Y}(0,-2) \neq f_X(0)f_Y(-2)$.

Problem 2:

A machine contains five components whose lifetime (in hours) has exponential distribution with mean 1000. The producer of the machine is considering to offer a guarantee that no more than two of the original components will have to be replaced during the first 1000 hours of use. Find the probability that such a guarantee will be violated in the following two cases:
(a) the components wear out independently and when a component does fail then the replacement used is of particularly high quality and will certainly last for the 1000 hours.
(b) the components wear out independently and no assumptions are made on the replacement components

Solution:

(a) Let X denote the lifetime of a single component. For each component, the probability that it wears out within the first 1000 hours of use is given by

$$p = \mathbb{P}(X < 1000) = \int_0^{1000} \frac{1}{1000} e^{-x/1000} dx = 0.6321.$$

Under the assumptions of item (a) the number of components, Y, that wear out within the first 1000 hours is binomially distributed $Y \sim Bin(5, p)$. We need to compute

$$\mathbb{P}(Y > 2) = \sum_{y=3}^{5} \mathbb{P}(Y = y) = 0.736$$

(b) If no assumptions are made on the failures of replacement components, then for any single original components the number of failures in 1000 hours has a Poisson distribution with mean 1. Since there are 5 components acting independently the total number of failures Y will have a Poisson distribution with mean 5. This implies

$$\mathbb{P}(Y > 2) = \sum_{y=3}^{\infty} \mathbb{P}(Y = y) = 1 - \sum_{y=0}^{2} \mathbb{P}(Y = y) = 0.875$$

Problem 3:

A printer receives, on average, 5 jobs per minutes. Assume the number of assigned printing jobs has a Poisson distribution. Find the probability:
(a) that in a 1 minute period no jobs are received
(b) that in a 2 minutes period fewer than 4 jobs are received
(c) that in 20 minutes no more than 102 jobs are received
(d) that out of five separate 1 minute periods there are exactly four in which 2 or more jobs are received.

Solution:
(a) Denoting by X the number of received jobs in a 1 minute period, we have that X has the Poisson distribution with mean 5, that is

$$\mathbb{P}(X = x) = e^{-5} \frac{5^x}{x!} \qquad x = 0, 1, 2, \ldots$$

Then $\mathbb{P}(X = 0) = e^{-5} = 0.007$
(b) The number of received jobs in a 2 minutes period, Y, has a Poisson distribution with mean 10. Hence

$$\mathbb{P}(Y < 4) = \sum_{y=0}^{3} \mathbb{P}(Y = y) = 0.010$$

(c) The number of received jobs in a 2 minutes period, W, has a Poisson distribution with mean $\mathbb{E}(W) = 100$ and variance $Var(W) = 100$. Since the mean is large we may use a normal approximation, that is W is well approximated by a guassian random variable $W' \sim N(100, 100)$. Then we find

$$\mathbb{P}(W \leq 102) \approx \mathbb{P}(W' \leq 102) = \mathbb{P}(Z < 0.2) = 0.5793$$

(d) In any 1 minute period chosen at random the probability of two or more jobs being received is $p = 1 - \mathbb{P}(X = 0) - p(X = 1) = 0.9596$. Then, using the binomial distribution $B(5, 0.9596)$, the probability that exactly four out of five 1-minute periods contain 2 or more jobs is $5(0.9596)^4(1 - 0.9596) = 0.171$

Problem 4:

The concentration C of a certain active agent in a pesticide must not exceed 12 part per million. The pesticide is made up in batches and in each batch the concentration is normally distributed, with mean 8 part per million and standard deviation 1.5 part per million.
(a) What proportion of batches exceeds the permitted maximum?

Two batches chosen at random and independently of each other are mixed so that the resulting concentration X is the average of the two concentration in the constituent batches: $X = \frac{1}{2}(C_1 + C_2)$.
(b) What is the probability that such a mixed sample has concentration value between 6 and 12 parts per million?
(c) What is the probability that the concentration of the mixed sample falls below 3 part per million (the level at which the pesticide cease to have any effect)?

Solution:
(a) Let C be the concentration of the active agent. Then $Z = \frac{C-8}{1.5} \sim N(0, 1)$. We have $\mathbb{P}(C > 12) = \mathbb{P}(Z > 8/3))$. From the Statistisch Compendium you see that $\mathbb{P}(Z > 8/3)) = 1 - p(Z < 8/3) = 1 - 0.9962 = 0.0038$. This is also the proportion of batches which exceed the maximum concentration.
(b) The concentration of the mixed batch will still be normally distributed being a linear combination of the two independent and identical gaussian random variables C_1 and C_2. We have $\mathbb{E}(X) = \frac{1}{2}(\mathbb{E}(C_1) + \mathbb{E}(C_2)) = 8$ and $Var(X) = \frac{1}{4}(Var(X_1) + Var(X_2)) = 1.125$ so that the standardized normal deviate of interest is $Z = \frac{X-8}{1.0607}$. We then have as before:

$$\mathbb{P}(6 < X < 10) = \mathbb{P}(-1.886 < Z < 3.771) = \mathbb{P}(Z < 3.771) - \mathbb{P}(Z < -1.886) = 0.9706$$

(c) $\mathbb{P}(X < 3) = \mathbb{P}(Z < -4.714) = 0.0000$ (corect to four decimal places)

Problem 5:

A person plays chess against a computer once a day. He can set the difficulty to 3 levels: easy, medium or difficult. The player sets an "easy" level with probability 1/2, a "medium" level with probability 1/4 and a "difficult" level with probability 1/4. If he plays an easy level game he has an 80% chance of winning; corresponding figures for medium and difficult level are 40% and 60% respectively.
(a) Find the probability that, on a given day, the player wins.
(b) If, on a particular day, the player has lost, which level has he most likely played?

Solution:
(a) Define the following events:
E: the level is set to easy
M: the level is set to medium
D: the level is set to difficult
W: the player win
Using the total law of probability we have

$$\mathbb{P}(W) = \mathbb{P}(E)\mathbb{P}(W|E) + \mathbb{P}(M)\mathbb{P}(W|M) + \mathbb{P}(D)\mathbb{P}(W|D)$$

Using the given information one finds $\mathbb{P}(W) = 13/20$.
(b) We have $\mathbb{P}(\bar{W}) = 1 - \mathbb{P}(W) = 7/20$. Applying Bayes formula three times we find:

$$\mathbb{P}(E|\bar{W}) = \frac{\mathbb{P}(E)\mathbb{P}(\bar{W}|E)}{\mathbb{P}(\bar{W})} = \frac{2}{7}$$

$$\mathbb{P}(M|\bar{W}) = \frac{\mathbb{P}(M)\mathbb{P}(\bar{W}|M)}{\mathbb{P}(\bar{W})} = \frac{3}{7}$$

$$\mathbb{P}(D|\bar{W}) = \frac{\mathbb{P}(D)\mathbb{P}(\bar{W}|D)}{\mathbb{P}(\bar{W})} = \frac{2}{7}$$

So it is most likely that he played with level of difficulty set to medium.

Problem 6:

Let $(X_1, X_2, X_3, X_4, X_5)$ be a random sample from a random variable X with expectation value $E(X) = \mu$ and variance $V(X) = \sigma^2$. Consider the following three estimators for the expected value μ

1. $T_1(X_1, X_2, X_3, X_4, X_5) = X_1$,
2. $T_2(X_1, X_2, X_3, X_4, X_5) = \frac{1}{3}(X_1 + X_3 + X_5)$,
3. $T_2(X_1, X_2, X_3, X_4, X_5) = \frac{1}{6}(X_1 + X_2 + X_3 + X_4 + X_5)$.

(a) Are the three estimators unbiased? Compute the bias.
(b) Compute the Mean Squared Error of the three estimators.
(c) Suppose $|\mu| < 2$ and $\sigma^2 = 1$. Which is the estimator with the smallest MSE ?

Solution:
(a) Bias:

$$\begin{aligned}
b(T_1) &:= E(T_1) - \mu = E(X_1) - \mu = \mu - \mu = 0, \\
b(T_2) &:= E(T_2) - \mu = E\left[\frac{1}{3}(X_1 + X_3 + X_5)\right] - \mu = \frac{1}{3}\left[E(X_1) + E(X_3) + E(X_5)\right] - \mu \\
&= \frac{1}{3}(\mu + \mu + \mu) - \mu = \mu - \mu = 0, \\
b(T_3) &:= E(T_3) - \mu = E\left[\frac{1}{6}(X_1 + X_2 + X_3 + X_4 + X_5)\right] - \mu \\
&= \frac{1}{6}(\mu + \mu + \mu + \mu + \mu) - \mu = -\frac{1}{6}\mu.
\end{aligned}$$

T_1 and T_2 are unbiased, T_3 is biased.
(b) For the MSE we use the formula $MSE(T) = V(T) + b(T)^2$. So we basically have to compute the variance

$$\begin{aligned}
V(T_1) &= V(X_1) = \sigma^2, \\
V(T_2) &= V\left[\frac{1}{3}(X_1 + X_3 + X_5)\right] = \frac{1}{9}\left[V(X_1) + V(X_3) + V(X_5)\right] \\
&= \frac{1}{9}(\sigma^2 + \sigma^2 + \sigma^2) = \frac{1}{3}\sigma^2, \\
V(T_3) &= E\left[\frac{1}{6}(X_1 + X_2 + X_3 + X_4 + X_5)\right] \\
&= \frac{1}{36}(\sigma^2 + \sigma^2 + \sigma^2 + \sigma^2 + \sigma^2) = \frac{5}{36}\sigma^2.
\end{aligned}$$

That means

$$\begin{aligned}
MSE(T_1) &= V(X_1) + b(T_1)^2 = \sigma^2, \\
MSE(T_2) &= V(T_2) + b(T_2)^2 = \frac{1}{3}\sigma^2, \\
MSE(T_3) &= V(T_3) + b(T_3)^2 = \frac{1}{36}(5\sigma^2 + \mu^2).
\end{aligned}$$

(c) For $|\mu| < 2$ and $\sigma^2 = 1$ we have

$$MSE(T_3) = \frac{1}{36}(5\sigma^2 + \mu^2) = \frac{5}{36} + \frac{1}{36}\mu^2 < \frac{9}{36} = \frac{1}{4}$$

and

$$MSE(T_1) > MSE(T_2) = \frac{1}{3}\sigma^2 = \frac{1}{3}.$$

Hence T_3 – although biased – has the lowest MSE.

Problem 7:

A pencil producer thinks that his pencil machine actually produces pencils of a mean length of $\mu = 17cm$. He assumes that the machine produces pencils with normally distributed lengths. The variance is unknown. He measures the lengths of five pencils from the production and obtains lengths of 19.2 cm, 17.4 cm, 18.5 cm, 16.5 cm, 18.9 cm.
(a) Compute sample mean $\overline{x}$ and sample variance s^2.
(b) Construct a 95 % - confidence interval for the mean pencil length. Does the result contradict the producer's opinion ?

Solution:

(a) We have sample mean $\overline{x} = 18.1$ and sample variance $s^2 = 1.265$.
(b) Since the variance is unknown we construct a t-confidence interval with $n - 1 = 4$ degrees of freedom. Thus

$$I = [\overline{x} - t_{n-1,\alpha/2} s/\sqrt{n}, \overline{x} + t_{n-1,\alpha/2} s/\sqrt{n}] = [18.1 - \sqrt{1.265}\, t_{4,0.025}/\sqrt{5}, 18.1 + \sqrt{1.265}\, t_{4,0.025}/\sqrt{5}]$$

With $t_{4,0.025} = 2.776$ that implies

$$I = [18.1 - 1.39, 18.1 + 1.39] = [16.71, 19.49].$$

Since $17 \in I$, these findings do not contradict the conjecture of the producer.

Problem 8:

Upgrading a certain software package requires installation of 68 new files. Files are installed consecutively. The installation time is random, but on the average, it takes 15 sec to install one file, with a variance of 11 sec^2
(a) What is the probability that the whole package is updated in less that 16 minutes?
(b) A new version of the package is released. It requires only N new files to be installed, and it is premised that 95% of the time upgrading takes less that 10 minutes. Given this information, compute N.

Solution:

(a) $n = 68$, $\mu = 15sec$ $Var(X) = 11sec^2$, $x = 16min = 960sec$

$$P(S_n \leq 960) \sim P[\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq \frac{960 - 1020}{\sqrt{748}}] = P[Z \leq -2.20] = 0.0143$$

(b) $N =?$, $\mu = 15sec$ $Var(X) = 11sec^2$, $x = 10min = 600sec$

$$P(S_n \leq 600) \sim P[\frac{S_n - N\mu}{\sigma\sqrt{N}} \leq \frac{600 - 15N}{\sqrt{11N}}] = P[Z \leq z] = 0.95$$

So for $N = 37$ you get the correct probability more than 95%.