MATH 2030 Elementary Probability/ Winter 2010

Test 3/ March 26, 2010

Solutions

Student Name: **ID-No.:**

You have 50 minutes to solve the following problems: Show your complete work. You can use a pre-written helpsheet with notes and formulas and a calculator: No other aids are permitted. The maximal score on this test, including bonus points, is 100.

Problem 1 **(20 points)**

Twelve percent of the population is left-handed. Approximate the probability that there are at least 20 left-handers in a school of 200 students. State your assumptions.

Denote:

X = number of left-handers in 200 students.

$X \sim Bin(200, 0.12)$

As $n \cdot p = 200 \cdot 0.12 \geq 10$ and $n \cdot (1-p) = 200 \cdot 0.88 \geq 10$, the normal approximation of the binomial distribution applies:

$$P(X \geq 20) \approx 1 - \Phi\left(\frac{20 - \frac{1}{2} - 200 \cdot 0.12}{(200 \cdot 0.12 \cdot 0.88)^{1/2}} \right)$$

$$= 1 - \Phi(-0.979)$$

$$= 1 - (1 - \Phi(0.979)) = \Phi(0.979) = \underline{\underline{0.8365}}$$

Problem 2: **(20 points)**

A wildlife toxicologist studying the effect of pollution on natural ecosystems measured the concentration of heavy metals in the blood of 25 Galapagos sea lions, Zalophus californianus. The sea lions sampled were all females and were 3 to 5 years old. Their mean concentration of heavy metals was 6.2 μg/ l: the standard deviation σ is known to be 1.5 μg/ l. Construct confidence intervals for the population mean

a) for a 95% significance level.

b) for a 99 % significance level.

Denote: $n = 25$ (sample size)

$\sigma = 1.5$ (standard deviation)

$\bar{X}_{25} = 6.2$ (sample mean)

a) $CI(95\%) = [\bar{X}_{25} - z_{0.975} \cdot \frac{\sigma}{\sqrt{n}}, \bar{X}_{25} + z_{0.975} \cdot \frac{\sigma}{\sqrt{n}}]$

where $z_{0.975}$ is the 97.5% percentile of the standard-normal distribution, (and $z_{0.975} = -z_{0.025}$); $z_{0.975} = 1.96$.

Therefore

μ = population mean $\in CI(95\%)$

$= [5.612, 6.788]$

with 95 % probability.

2) b) 99 % confidence interval for μ:

$$z_{0.995} = 2.575 \quad , \quad z_{0.005} = -2.575$$

$$\Rightarrow \mu \in CI(99\%)$$

$$= \left[\bar{X}_{25} - 2.575 \cdot \frac{1.5}{\sqrt{25}} \, , \, \bar{X}_{25} + 2.575 \cdot \frac{1.5}{\sqrt{25}} \right]$$

$$= [5.428 \, , \, 6.973]$$

Problem 3: **(20 points)**

The time (in hours) required to repair a machine is an exponentially distributed random variable with parameter $\lambda = 1/2$. What is

a) the probability that a repair time exceeds 2 hours?

b) the conditional probability that a repair takes at least 10 hours, given that its duration exceeds 9 hours?

Denote T: repair time

$$T \sim Exp(\tfrac{1}{2})$$

a) $P(T > 2) = e^{-\frac{1}{2} \cdot 2} = e^{-1} = \underline{\underline{0.368}}$

by using the survival function of the exponential distribution.

b) $P(T \geq 10 \mid T > 9)$

$= P(T > 1)$ because of the memory-less property of the exponential distribution

$= e^{-\frac{1}{2} \cdot 1} = \underline{\underline{0.607}}$

Problem 4: **(20 points)**

A box contains four tickets, numbered 0, 1, 1, and 2. Let S_n be the sume of the numbers obtained from n draws at random with replacement from the box.

a) Display the distribution of S_2 in a suitable table.

b) Find $P(S_{50} = 50)$ approximately.

c) Find an exact formula for $P(S_n = k)$ (k= 0,1, 2, ...).

a)

x	$\{S_2 = x\}$	$P(S_2 = x)$
0	$\{(0,0)\}$	$\frac{1}{4}\cdot\frac{1}{4} = \frac{1}{16}$
1	$\{(0,1),(1,0)\}$	$\frac{1}{4}\cdot\frac{1}{2} + \frac{1}{2}\cdot\frac{1}{4} = \frac{1}{4}$
2	$\{(0,2),(1,1),(2,0)\}$	$\frac{1}{4}\cdot\frac{1}{4} + \frac{1}{2}\cdot\frac{1}{2} + \frac{1}{4}\cdot\frac{1}{4} = \frac{3}{8}$
3	$\{(1,2),(2,1)\}$	$\frac{1}{2}\cdot\frac{1}{4} + \frac{1}{4}\cdot\frac{1}{2} = \frac{1}{4}$
4	$\{(2,2)\}$	$\frac{1}{4}\cdot\frac{1}{4} = \frac{1}{16}$

e.g. $P(S_2 = 1) = P(X_1 = 0 \text{ and } X_2 = 1) + P(X_1 = 1 \text{ and } X_2 = 0)$ where

X_1 and X_2 denote the outcomes of the first and the second draw, resp.;

$$P(S_2 = 1) = P(X_1 = 0)\cdot P(X_2 = 1) + P(X_1 = 1)\cdot P(X_2 = 0)$$

$$= \frac{1}{4}\cdot\frac{2}{4} + \frac{2}{4}\cdot\frac{1}{4} = \frac{1}{4}.$$

4) b) $P(S_{50} = 50)$

$= P(S_{50} \leq 50) - P(S_{50} \leq 49)$

These probabilities can be calculated approximatively by using the Central Limit Theorem:

$$P(S_{50} \leq 50) = P\Big(\underbrace{\frac{S_{50} - 50 \cdot E(X)}{\sqrt{50} \cdot \sigma_X}}_{\approx Z \sim N(0,1)} \leq \frac{50 - 50 \cdot E(X)}{\sqrt{50} \cdot \sigma_X}\Big)$$

Here $E(X) = \frac{1}{4} \cdot 0 + \frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 2 = 1$ and

$$E(X^2) = \frac{1}{4} \cdot 0^2 + \frac{1}{2} \cdot 1^2 + \frac{1}{4} \cdot 2^2 = 1.5$$

$\Rightarrow$ $Var(X) = E(X^2) - (E(X))^2 = 0.5$ and

$\sigma_X = \sqrt{Var(X)} = 0.707$.

We obtain $P(S_{50} \leq 50) \approx \Phi(0) = \frac{1}{2}$

and

$$P(S_{50} \leq 49) = P\Big(\frac{S_{50} - 50 \cdot 1}{\sqrt{50} \cdot 0.707} \leq \frac{49 - 50 \cdot 1}{\sqrt{50} \cdot 0.707}\Big)$$

$$\approx \Phi(-0.2) = 1 - \Phi(0.2) = 1 - 0.5793$$
$$= 0.4207$$

4) b) cont'd.:

$$P(S_{50} = 50) \approx 0.5 - 0.4207 = \underline{\underline{0.0793}}$$

4) c)

$$P(S_n = k) = \begin{cases} \displaystyle\sum_{\substack{n_0 + n_1 + 2n_2 = k \\ n_0 + n_1 + n_2 = n}} \left(\tfrac{1}{4}\right)^{n_0} \cdot \left(\tfrac{1}{2}\right)^{n_1} \cdot \left(\tfrac{1}{4}\right)^{n_2} & \text{for all } 0 \le k \le 2n \\ 0 & \text{for } k > 2n \end{cases}$$

Problem 5) **(20 points)**

If you buy a lottery ticket in 50 successive lotteries, in each of which your chance of winning the price is 1/100, what is the (approximate) probability that you will win the a prize

a) at least once?

b) exactly once?

c) three or more times?

Use the Poisson-distribution: For

X = # wins in 50 successive lotteries

$X \sim \text{Poisson}(\lambda)$ where $\lambda = 50 \cdot \frac{1}{100} = 0.5$.

a) $$P(X \geq 1) = 1 - P(X=0) = 1 - e^{-0.5} = \underline{\underline{0.3935}}$$

b) $$P(X = 1) = e^{-0.5} \cdot \frac{0.5^1}{1!} = \underline{\underline{0.303}}$$

c) $$P(X \geq 3) = 1 - P(X \leq 2) = 1 - \left(e^{-0.5} + e^{-0.5} \cdot \frac{0.5^1}{1!} + e^{-0.5} \cdot \frac{0.5^2}{2!}\right) = \underline{\underline{0.0144}}$$

Bonus problem: **(10 points)**

The number of eggs laid on a tree leaf by an insect of a certain type is a Poisson random variable with parameter λ. However, such a random variable can be observed only if it is positive, since if it is 0 then we cannot know that such an insect was on the leaf. If we let Y denote the observed number of eggs, then

P(Y = i) = P(X = i | X > 0);

where X is Poisson-distributed with parameter λ. Find E(Y).

$$E(Y) = \sum_{i=0}^{\infty} i \cdot P(Y=i) = \sum_{i=1}^{\infty} i \cdot P(Y=i) ;$$

here $P(Y=i)$, for any $i \geq 1$, is given by

$$P(X=i \mid X>0) = \frac{P(X=i \text{ and } X>0)}{P(X>0)}$$

$$= \frac{P(X=i)}{P(X>0)} = \frac{P(X=i)}{1-P(X=0)}$$

$$= \frac{e^{-\lambda} \cdot \lambda^i / i!}{1-e^{-\lambda}} .$$

Therefore,

$$E(Y) = \sum_{i=1}^{\infty} i \cdot \frac{e^{-\lambda} \cdot \lambda^i / i!}{1-e^{-\lambda}} = \frac{1}{1-e^{-\lambda}} \cdot e^{-\lambda} \cdot \sum_{i=1}^{\infty} \frac{\lambda^i}{(i-1)!}$$

$$= \frac{1}{1-e^{-\lambda}} \cdot e^{-\lambda} \cdot \lambda \cdot \sum_{i=0}^{\infty} \frac{\lambda^i}{i!}$$; the latter series

Bonus Problem cont'd:

is equal to e^{λ} (Taylor series),

and this yields:

$$E(Y) = \frac{\lambda}{1-e^{-\lambda}}$$

E.g. for $\lambda = 3$: $E(Y) = \frac{3}{1-e^{-3}} = \underline{\underline{3.157}}$

$E(Y)$ represents the average observed number of eggs and includes a "positive ascertainment bias", i.e. a bias that is due to the non-reporting of negative observations.