Econ 4130 Exam 2021 H - POSTPONED

Problem 1

Consider a sequence of Bernoulli trials[1] with constant success probability, *p*, where the outcomes from different trials are independent events. Let X be the number of trials until and including the first "success" (S). Then *X* is geometrically distributed with success probability *p* (written shortly: $X \sim \text{geom}(p)$). The probability mass function (*pmf*) of *X* is

$$f(x) = P(X = x) = (1-p)^{x-1} p \quad \text{for } x = 1, 2, 3, \ldots$$

where $0 < p < 1$ is a parameter.

A. **i)** Show that $P(X \le 2) = 2p - p^2$.

ii) Derive the cumulative distribution function (*cdf*) of *X*.

[**Hint.** Note that the event $(X > x)$ is equivalent with the event that the first *x* trials result in failures.]

iii) Show that the geom(p) distribution "lacks memory" in the sense that

$$P(X > n + k \mid X > n) = P(X > k)$$

where *n, k* are natural numbers.

B. It turns out that the moment generating function (*mgf*) of *X* is

$$M(t) = \frac{p}{e^{-t} - 1 + p} \quad \text{(well defined for } t < -\ln(1-p) \text{)}$$

Use the mgf to show that $E(X) = \frac{1}{p}$ and $\text{var}(X) = \frac{1-p}{p^2}$.

C. Now, for the same sequence of Bernoulli trials determining *X*, let *Y* be the number of failures *before* the first success. In other words, $Y = X - 1$.

i) Find the pmf of *Y*, and write up $E(Y)$ and $\text{var}(Y)$.

ii) Find the mgf of *Y*.

---

[1] Each trial has one of two possible outcomes, "success" (S) or "failure" (F). The probability of success is $p = P(S)$.

D. Show that the mgf of X is $M_X(t) = \dfrac{pe^t}{1-e^t(1-p)} = \dfrac{p}{e^{-t}-1+p}$, and that the mgf is well defined for $t < -\ln(1-p)$.

[Hint. You are reminded of a geometric series,
$\sum_{k=0}^{\infty} a^k = a^0 + a^1 + a^2 + a^3 + \cdots = \dfrac{1}{1-a}$ whenever $|a| < 1$. In addition, a common factor in a series may be taken outside the series, or, in symbols,
$\sum_{k=0}^{\infty}\left(ca^k\right) = c + ca + ca^2 + \cdots = c\sum_{k=0}^{\infty} a^k = c\left(1 + a + a^2 + \cdots\right)$.
Note also that $e^{tx} = e^t e^{t(x-1)}$. **]**

Problem 2

Introduction. Many casinos have *slot machines* that are a special type of gambling machines. A trial on such a machine consists of the player inserting a coin in the machine, then pushes a button, and watches figures on a screen spin until they stop in a random pattern. Some patterns generate a win (considered as "success") consisting of a small or larger number of coins coming out of the machine to the player. Most patterns, however, will not lead to any win at all and the money invested (inserted) at the start of the trial is lost. Such an outcome we consider a "failure".

We assume that the outcomes of different trials are independent and that the probability, *p*, of a win (success) is the same in each trial. Hence, the trials can be considered as Bernoulli trials with respect to the two outcomes, win or not-win.

As in **problem 1**, let X be the number of trials until (and including) the first win. Then $X \sim \text{geom}(p)$ distributed with pmf

$$f(x) = P(X = x) = (1-p)^{x-1} p \quad \text{for } x = 1, 2, 3, \ldots$$

Data. A player has observed $n = 30$ observations of X, denoted by $x_1, x_2, \ldots, x_{30}$, on a single slot machine. The data are presented in **table 1**. For example, $x_2 = 7$ means that for the second observation it took 6 trials with not-wins before a win occurred in the 7th trial.

Table 1. 30 independent observations of *X* on a single slot machine.

2	7	7	8	5	15	3	7	2	9	1	1	5	8	2
12	7	8	8	4	6	21	1	4	4	5	1	4	10	7

Sum: $\sum_{i=1}^{30} x_i = 184$

Model. We assume that the data are observations of random variables (rv), $X_1, X_2, \ldots, X_n$ (with $n = 30$), that are independent and identically distributed (*iid*) and where X_i has the same distribution as *X*.

In some of the following questions you may need information from **problem 1**.

A. **i)** Explain why $\tilde{p} = \frac{1}{\bar{X}}$, where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, is the moment method estimator (*mme*) of *p*. Calculate the estimate based on the data in **table 1**.

ii) Put $\mu = E(X_i) = \frac{1}{p}$ and show that $\text{var}(\bar{X}) = \frac{\mu^2 - \mu}{n}$.

[Hint. You may consider as known the fact that

$\text{var}(X_i) = \sigma^2 \;\Rightarrow\; \text{var}(\bar{X}) = \frac{\sigma^2}{n}$ **]**

B. **i)** Use the central limit theorem (CLT) and (if needed) Slutsky's lemma to show that $\frac{\bar{X} - \mu}{\sqrt{\bar{X}^2 - \bar{X}}}\sqrt{n}$ converges in distribution to a standard normal rv, $Z \sim N(0,1)$, as $n \to \infty$.

ii) Use the result in **i)** to develop an approximate 95% confidence interval (CI) for μ and transform this CI into a CI for *p* with the same degree of confidence.

iii) Calculate the CI for *p* developed in **ii)** based on the data in **table 1**.

C. **i)** Show that the maximum likelihood estimator (*mle*), $\hat{p}$, for *p* is equal to the mme, $\tilde{p}$.

ii) Show that the Fisher information for one observation is

$$I(p) = \frac{1}{p^2(1-p)}$$

D. **i)** Using the fact that the conditions for "good behaviour" of mle estimators are fulfilled in this case, find an approximate 95% CI for *p* based on the mle $\hat{p}$. Calculate the CI based on the data in **table 1** and compare with the CI in **Biii)**.

ii) Write up a test statistic for testing $H_0: p = 0.1$ versus $H_1: p \neq 0.1$. Perform the test using the level of significance approximately 5% and interpret the result.

E. The player suspects that the probability of a win may vary between slot machines. In addition to the slot machine used in **table 1** (called machine 1), the player collects similar observations from two more slot machines (machine 2 and 3), with 20 observations from each.

Let $n_j, p_j \quad j=1,2,3$ be the sample sizes and the probabilities of a win respectively from the three machines ($n_1=30,\ n_2=n_3=20$).

Let $X_{ij}, \quad i=1,2,\ldots,n_j, \quad j=1,2,3$ be the rv's behind the observations. We assume that all X_{ij} are independent where $X_{ij} \sim \text{geom}(p_j), \quad i=1,2,\ldots,n_j, \quad j=1,2,3$.

We want to test $H_0: p_1 = p_2 = p_3$ versus H_1: "*at least two* p_j *different*"
using a likelihood ratio test (LR test).

i) Let $l_j(p_j)$ be the log likelihood for the data from machine *j* only. Explain why the log likelihood for the complete data set under the full model is $l_{full}(p_1, p_2, p_3) = l_1(p_1) + l_2(p_2) + l_3(p_3)$, and why the mle $\hat{p}_j$ must be as in **Ci)** depending on data from machine *j* only (ie. $\hat{p}_j = 1/\bar{X}_j$, where $\bar{X}_j$ is the mean from machine *j*).

The complete data are summarized in **table 2.**

Table 2

Slot machine j	n_j	$\sum_{i=1}^{n_j} x_{ij}$	$\hat{p}_j$	$l_j(\hat{p}_j)$
1	30	184	0.1630	-81.8216
2	20	170	0.1176	-61.5758
3	20	190	0.1053	-63.9342
sum	70	544		-207.332

ii) **Reduced model.** Call the common value under H_0 of p_1, p_2, p_3 for *p*. Explain briefly why the mle under H_0 must be $\hat{p} = \dfrac{n_1+n_2+n_3}{\sum_{j=1}^{3}\sum_{i=1}^{n_j} X_{ij}}$.

Calculate the observed value of the log likelihood under H_0 (that we call $l_{red}(p)$) replacing *p* by $\hat{p}$, ie., $l_{red}(\hat{p})$.

iii) Test H_0 by the LR test on approximately 5% level of significance and comment on the result.