Homework #6 Adam Rurth

1) Question 1. (20 pts.) An employee of a bank wants to test the null hypothesis that on the average the ban cashes 10 bad checks per day against the alternative that this figure is too small. If he takes a random sample (from the population having a Poisson distribution with the parameter λ) and decides to reject the null hypothesis if and only if the mean of the sample exceeds 12.5, what decision will he make if he gets x¯ = 11.2 and determine if it will be in error when
(a) λ = 11.5 mm, (b) λ = 10.0 mm.

Test: null whether average bank checks 10 bad checks per day $H_0: \lambda_0 = 10$
vs. alternative figure is too small $H_1: \lambda_1 > 10$

$H_0: \lambda_0 = 10,\ H_1: \lambda_1 > 10$
$X_1, \dots X_n \sim \text{Poisson}(\lambda)$, $\bar{x}$ estimates λ
reject iff $\bar{x} > 12.5$
a) $\lambda = 11.5$ he will not reject the null, since $\lambda = 11.5 \not> 12.5$ which was our decision rule.
We have not rejected the null, thus a type II error will occur.

b) $\lambda = 10$ this saying that the sample average (λ) is exactly equal to 10 which matches the null hypothesis thus, no error will occur.

2) Question 2. (20 pts.) A single observation of a random variable having a geometric distribution is used to test the null hypothesis θ = θ0 against the alternative hypothesis θ = θ1 > θ0. If the null hypothesis is rejected if and only if the observed value of the random variable is greater than or equal to the positive integer k, find expressions for the probabilities of type I and type II errors.

null rejected $\Leftrightarrow X \geq k : k \in \mathbb{Z}^+$
we know $X \sim \text{Geometric}(\theta)$
PMF: $P(X=k) = \theta(1-\theta)^{k-1}$, cdf: $P(X \le k) = 1-(1-\theta)^k$
$H_0: \theta = \theta_0,\ H_1: \theta_1 > \theta_0$

a) Type I error: $P(\text{Type I error}) = P(H_0 \text{ rejected} \mid H_0 \text{ is true})$
want to find α, the significance level when r.v. X is $\geq k$ under θ_0

$$\alpha = P(X \geq k \mid \theta = \theta_0) = \sum_{i=k}^{\infty} (1-\theta_0)^{i-1}\theta_0$$

this series is the tail of Geometric dist starting at k,

$$\alpha = (1-\theta_0)^{k-1}$$
$X \geq k$

$= P(H_0 \text{ rejected} \mid H_0 \text{ is true})$
$= P(X \geq k \mid \theta = \theta_0)$
$= 1 - P(X \leq k-1 \mid \theta = \theta_0)$
$= 1 - [1-(1-\theta_0)^{k-1}]$
$= (1-\theta)^{k-1}$

b) Likewise, we know type II error is when $P(H_0 \text{ not rejected} \mid H_0 \text{ is false})$

$P(\text{Type II error}) = P(H_0 \text{ not rejected} \mid H_0 \text{ is false})$
probability of $X < k$ under H_1 (ie θ_1)

$$\beta = P(X < k \mid \theta = \theta_1) = \sum_{i=1}^{k-1} (1-\theta_1)^{i-1}\theta_1$$

this sum is the cdf of a Geometric distribution up to $k-1$

$$\beta = 1-(1-\theta_1)^{k-1}$$

3) Question 3. (20 pts.) A single observation of a random variable having a uniform distribution with α = 0 us used to test the null hypothesis β = β0 against the alternative hypothesis β = β0 + 2. If the null hypothesis is rejected if and only if the random variable takes on a value greater than β0 + 1, find the probabilities of type I and type II errors.

$\text{Unif}[\alpha, \beta]$, $\alpha = 0$,
know pdf since $\alpha = 0$, $f(x) = \begin{cases} \frac{1}{\beta} & : x \in [0, \beta] \\ 0 & \text{elsewhere} \end{cases}$

$H_0: \beta = \beta_0$ $H_1: \beta = \beta_0 + 2$
If rejected $\Leftrightarrow X > \beta_0 + 1$
a) $P(\text{Type I}) = P(\text{reject } H_0 \mid H_0 \text{ is true})$

$$\alpha = P(X > \beta_0 + 1 \mid \beta = \beta_0) = 0$$

since X is uniformly distributed from 0 to β
the probability that X exceeds $\beta_0 + 1$ is 0,
since $\beta_0 + 1$ is outside the range of X

b) $P(\text{Type II error}) = P(\text{not reject } H_0 \mid H_0 \text{ is false})$

$$\beta = P(X \leq \beta_0 + 1 \mid \beta = \beta_0 + 2) = \frac{\beta_0 + 1}{\beta_0 + 2}$$

since probability of X doesn't exceed $\beta_0 + 1$ when $\beta = \beta_0 + 2$
and $X \sim \text{Unif}[\alpha, \beta_0 + 2]$ this means that the probability of
$X \geq \beta_0 + 1$ is $\frac{\beta_0 + 1}{\beta_0 + 2}$
ratio of the lengths of the supports $[0, \beta_0 + 1]$ and $[0, \beta_0 + 2]$

4) Question 4. (20 pts.) Let X1 and X2 constitute a random sample of size 2 from the population given by $f(x) = \begin{cases} \theta x^{\theta - 1} & : 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$
If the critical region x1x2 ≥ 34 is used to test the null hypothesis θ = 1 against the alternative hypothesis θ = 2, what is the power of this test at θ = 2?

$X_1, X_2 \sim f_\theta$ critical region $X_1 X_2 \geq \frac{3}{4}$
$H_0: \theta = 1,\ H_1: \theta = 2$
want to find power of the test at $\theta = 2$
calculate joint pdf: $f(x_1, x_2; \theta = 2) = f(x_1; \theta) f(x_2; \theta)$ assuming independence
$\Rightarrow \theta x_1^{\theta-1} (\theta x_2^{\theta-1}) = \theta^2 x_1^{\theta-1} x_2^{\theta-1} = (2)^2 x_1^{2-1} x_2^{2-1} = 4 x_1 x_2$
$f(x_1, x_2; \theta = 2) = 4 x_1 x_2$; $x_1, x_2 \in (0,1)$
given the critical region $x_1 x_2 \geq \frac{3}{4}$ we know
$x_2 \geq \frac{3}{4x_1}$ $\because x_2 > 0$ and bounded above by 1 $\left(\frac{3}{4x_1} \leq x_2 \leq 1\right)$
since $x_2 \leq 1$ this means that $x_1 x_2$ can only be as large as x_1
we know that x_1 must be at least $\frac{3}{4}$ as large therefore $x_1 x_2 \geq \frac{3}{4}$ $\left(\frac{3}{4} \leq x_1 \leq 1\right)$

$$\text{power} = \int_{\frac{3}{4}}^{1} \int_{\frac{3}{4x_1}}^{1} 4x_1 x_2 \, dx_1 dx_2 \rightarrow \int_{\frac{3}{4x_1}}^{1} 4x_1 x_2 \, dx_2 = 4x_1 \int_{\frac{3}{4x_1}}^{1} x_2 \, dx_2 \rightarrow \int_{\frac{3}{4}}^{1} 2x_1 - \frac{9}{8x_1} \, dx_1 \approx 0.113858$$

$$= 4x_1 \left[\frac{x_2^2}{2}\right]_{\frac{3}{4x_1}}^{1}$$

$$= 4x_1 \left(\frac{1}{2} - \frac{1}{2}\left(\frac{3^2}{4x_1}\right)\right)$$

$$= 2x_1 - \frac{9}{8x_1}$$

Thus, the power of the test at $\theta = 2$, is $0.113858 \approx 0.114$

5)

Question 5. (20 pts.) Given a random sample of size n from a normal population with μ = 0, use the Neyman–Pearson lemma to construct the most powerful critical region of size α to test the null hypothesis σ = σ0 against the alternative σ = σ1 (σ1 < σ0).

$x_1, \dots x_n \overset{iid}{\sim} N(\mu=0, \sigma)$

Find most powerful region size of a test : $H_0: \sigma=\sigma_0$, $H_1: \sigma=\sigma_1$

Find likelihoods for H_0, H_1 :

we know $f(x;\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\left(\frac{x_i}{2\sigma}\right)^2}$

$$L(\sigma|x) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\left(\frac{x_i}{2\sigma}\right)^2} = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left[\frac{-1}{2}\sum x_i^2\right] \quad (1)$$

Specifically:

$$L_0 = L(\sigma_0) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\left(\frac{x_i}{2\sigma_0}\right)^2} = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left[\frac{-1}{2}\sum_{i=1}^{n} x_i^2\right]$$

$$L_1 = L(\sigma_1) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\left(\frac{x_i}{2\sigma_1}\right)^2}$$

$$\Rightarrow \frac{L_1}{L_0} = \frac{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\left(\frac{x_i}{2\sigma_1}\right)^2}}{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\left(\frac{x_i}{2\sigma_0}\right)^2}} = \frac{\cancel{\left(\frac{1}{\sqrt{2\pi}}\right)^n} \prod_{i=1}^{n} \frac{1}{\sigma_1} \exp\left(\frac{-x_i}{2\sigma_1}\right)^2}{\cancel{\left(\frac{1}{\sqrt{2\pi}}\right)^n} \prod_{i=1}^{n} \frac{1}{\sigma_0} \exp\left(-\frac{x_i}{2\sigma_0}\right)^2}$$

rearrange

$$= \left(\frac{\sigma_0}{\sigma_1}\right)^n \exp\left[\sum_{i=1}^{n}\left(\frac{x_i^2}{2\sigma_1^2} - \frac{x_i^2}{2\sigma_0^2}\right)\right]$$

$$= \left(\frac{\sigma_0}{\sigma_1}\right)^n \exp\left[\sum_{i=1}^{n} x_i^2\left(\frac{1}{2\sigma_1^2} - \frac{1}{2\sigma_0^2}\right)\right]$$

$$= \left(\frac{\sigma_0}{\sigma_1}\right)^n \exp\left[\left(\frac{1}{2\sigma_1^2} - \frac{1}{2\sigma_0^2}\right)\sum_{i=1}^{n} x_i^2\right]$$

Using Neyman-Pearson Lemma, we must conduct the likelihood ratio test for $H_0: \sigma=\sigma_0$, vs $H_1: \sigma=\sigma_1$ $(\sigma_1 > \sigma_0)$

We have the most powerful region of a given size α, $\Lambda(x) = \frac{L(\sigma_1|x)}{L(\sigma_0|x)}$

$$L(\sigma|x) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left[\frac{-1}{2\sigma^2}\sum_{i=1}^{n} x_i^2\right] \text{ by equation (1)}$$

Thus the likelihood ratio becomes (from the calculations above)

$$\Lambda(x) = \frac{L(\sigma_1|x)}{L(\sigma_0|x)} = \left(\frac{\sigma_0}{\sigma_1}\right)^n \exp\left[\frac{-1}{2}\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right)\sum_{i=1}^{n} x_i^2\right]$$

Simplifying, we know under H_1; $\sigma_1 > \sigma_0$

The difference $\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2} > 0 \Leftrightarrow \frac{1}{\sigma_1^2} > \frac{1}{\sigma_0^2}$

$\Lambda(x)$ is a decreasing function of the test statistic $T(x) = \sum x_i^2 \sim \chi^2(n)$

follows Chi-Square under H_0

Namely inside the exponent, $\underbrace{-\frac{1}{2}\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right)}_{>0} \underbrace{\sum_{i=1}^{n} x_i^2}_{T(x) \sim \chi^2(n)}$

Thus likelihood decreases as $\sum x_i^2$ increases

$\therefore$ $\Lambda(x)$ is a monotonically decreasing function of $T(x) = \sum x_i^2$

Since we know this about $\Lambda(x)$, the most powerful test rejects H_0 when $\Lambda(x)$ is below some threshold, k

$$P(T(x) < k \mid H_0) = \alpha$$

Find k by looking up the α-quantile of χ^2 distribution with n degrees of freedom.

$$C = \left\{x \mid T(x) = \sum_{i=1}^{n} < k\right\}$$ most powerful critical region.