Some of my white board writing from week 10

I discussed the generalized likelihood ratio test for two examples related to the normal distribution:

(a) Testing $H_0: \mu = \mu_0$ vs $H_1: \mu \neq \mu_0$ for a sample of n i.i.d. $N(\mu, \sigma^2)$ (σ^2 assumed known)

In this case

$$-2\left[\ln L(\mathbf{X}, H_0) - \ln L(\mathbf{X}, \bar{x})\right] = -2\left[-\frac{\sum_{i=1}^n (X_i-\mu_0)^2}{2\sigma^2} + \frac{\sum_{i=1}^n (X_i-\bar{X})^2}{2\sigma^2}\right] =$$

$$= \frac{1}{\sigma^2}\left[\sum_{i=1}^n (X_i-\mu_0)^2 - \sum_{i=1}^n (X_i-\bar{X})^2\right] = \frac{n(\bar{X}-\mu_0)^2}{\sigma^2}$$

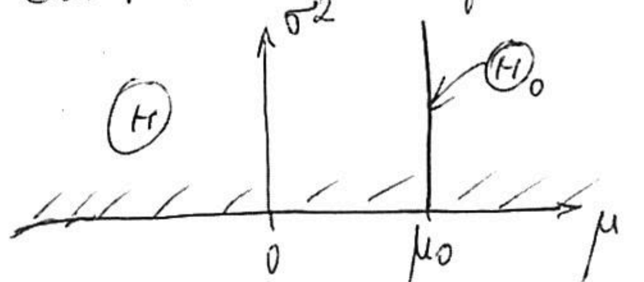
This should be $\sim \chi^2_1$ asymptotically but in this case, because of dealing with normal, the result is precise (not only asymptotic). Indeed, we know that under H_0: $\bar{X} \sim N(\mu_0, \frac{\sigma^2}{n}) \Rightarrow \frac{\sqrt{n}(\bar{X}-\mu_0)}{\sigma} \sim N(0,1) \Rightarrow \frac{n(\bar{X}-\mu_0)^2}{\sigma^2} \sim \chi^2_1$

The GLRT is: $\varphi^* = \begin{cases} 1 & \text{if } \frac{n(\bar{X}-\mu_0)^2}{\sigma^2} > \chi^2_{\alpha,1} \\ 0 & \text{if } \frac{n(\bar{X}-\mu_0)^2}{\sigma^2} \le \chi^2_{\alpha,1} \end{cases}$ and

is equivalent to the standard Z-test in this case.

(b) Testing $H_0: \mu = \mu_0$ vs $H_1: \mu \neq \mu_0$ again but when σ^2 is unknown. Hence we are testing in effect:

$H_0: \begin{cases} \mu = \mu_0 \\ \sigma^2 > 0 \end{cases}$ vs $H_1: \begin{cases} \mu \neq \mu_0 \\ \sigma^2 > 0. \end{cases}$ In terms of the notation of GLRT: we have a parameter vector $\theta = \binom{\mu}{\sigma^2}$ and



Θ_0 is one-dimensional subspace as sketched, whereas Θ is "anything" above the Ox axis.

The dimensions K, r, s, as discussed in Section 6.10.1 of Lecture 6, p. 56, are $K=2$, $r=1$, $s=1$.
To perform the GLRT we need to maximize $L(\mathbf{X}, \theta)$ under the Hypothesis and under the alternative.
i) under the hypothesis: $\mu = \mu_0$, so we need to optimize w.r. to σ only. $\ln L = -\frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu_0)^2 + \text{const}$

$$\frac{\partial}{\partial\sigma^2}\ln L = 0 = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(X_i-\mu_0)^2 \quad \text{implies}$$

$$\hat{\sigma}^2_{H_0} = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu_0)^2$$

and $$\sup L|H_0 = \frac{1}{(\sqrt{2\pi})^n(\hat{\sigma}^2_{H_0})^{n/2}}\exp\left(-\frac{n}{2}\right)$$

ii) without the restriction of H_0, we have to maximise $\ln L$ w.r. to both μ and σ^2, i.e. solve the system

$$\frac{\partial}{\partial\mu}\ln L = 0, \qquad \frac{\partial}{\partial\sigma^2}\ln L = 0$$

This leads to $\hat{\mu} = \bar{X}$, $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$

When plugging-in, we get the $\sup L$ without any restriction and it is $$\sup L = \frac{1}{(\sqrt{2\pi})^n(\hat{\sigma}^2)^{\frac{n}{2}}}\exp\left(-\frac{n}{2}\right)$$

Hence $$-2\log\Lambda = -2\log\left(\frac{\hat{\sigma}^2}{\hat{\sigma}^2_{H_0}}\right)^{\frac{n}{2}} = n\log\left(\frac{\hat{\sigma}^2_{H_0}}{\hat{\sigma}^2}\right)$$ and the GLRT is

$$\varphi^* = \begin{cases} 1 & \text{if } n\log\left(\frac{\hat{\sigma}^2_{H_0}}{\hat{\sigma}^2}\right) > \chi^2_{\alpha,1} \\ 0 & \text{if } n\log\left(\frac{\hat{\sigma}^2_{H_0}}{\hat{\sigma}^2}\right) \le \chi^2_{\alpha,1} \end{cases}$$

(the degrees of freedom are $=1$ since in this case $r = 1 = K - S$ $(K=2, S=1)$.
Note that now the convergence of $-2\log\Lambda$ to the limiting χ^2_1 is only <u>asymptotic</u> (not precise as in case a)).
But $-2\log\Lambda = n\log\left(1 + \frac{(\bar{X}-\mu_0)^2}{\hat{\sigma}^2}\right) \approx \frac{n(\bar{X}-\mu_0)^2}{S^2}$, so it is <u>"almost" equivalent to the standard t-test for $\mu=\mu_0$ when σ^2 is unknown.</u>

Continuation:

Some of my white board writing in week 10 ③

1.) Regarding the multinomial distribution:

I explained the formula

$$P(X_1=x_1, X_2=x_2, \dots, X_k=x_k) = \frac{n!}{x_1!\,x_2!\cdots x_k!} w_1^{x_1} w_2^{x_2} \cdots w_k^{x_k},$$

$0<w_i<1$, $\sum_{i=1}^{k} w_i = 1$ for calculating the probability of a particular outcome $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{pmatrix}$ with $x_1+x_2+\dots+x_k=n$, n being the number of independent trials.

I also discussed two examples:

i) If a die is tossed 6 times, what is the probability that each number $(1,2,3,4,5,6)$ turns up once.
Applying the above formula with $k=6$, $x_i=1$, $i=1,2,\dots,6$, and $w_i=\frac{1}{6}$, $i=1,2,\dots,6$ we get $6!\left(\frac{1}{6}\right)^6 = \frac{5}{324}$

ii) Out of 7 tosses, what is the probability that each number $(1,2,3,4,5,6)$ turns up at least once.

Answer: $$6\cdot\frac{7!}{2!\,(1!)^6}\cdot\left(\frac{1}{6}\right)^7 = \frac{35}{648}$$

2.) I discussed in detail the proof of Theorem 7.2. (p. 60) but I see that all details are presented in the lecture note so I will abstain from reproducing them again here.

3.) I discussed a simple method to derive the density of the r-th order statistic as stated in Theorem 7.3, p. 61

Details: we introduce a <u>discrete</u> random variable
$Y = \{$ number of realisations $X_1, X_2, \dots, X_n$ that happen to be $\le x \}$
Then $Y \sim Bin(n, F_X(x))$.
Now we first derive $F_{X_{(n)}}(x)$ (the cdf of $X_{(n)}$) and then differentiate it to find the density. The <u>main</u> observation we make is that

$$F_{X_{(r)}}(x) = P(X_{(r)} \le x) = P(Y \ge r)$$

Hence we can state that

$$F_{X_{(r)}}(x) = \sum_{k=r}^{n} \binom{n}{k} \left(F_X(x)\right)^k \left(1 - F_X(x)\right)^{n-k}$$

Now to get the density we need to differentiate the each of ~~the~~ the summands in $\sum\limits_{k=r}^{n}$ by applying the $(uv)' = u'v + v'u$ formula each time.
We get:

$$\begin{aligned} f_{X_{(r)}}(x) = & \binom{n}{r} r f_X(x) F_X(x)^{r-1} (1 - F_X(x))^{n-r} \cancel{- \binom{n}{r}(n-r) F_X(x)^r (1-F_X(x))^{n-r-1} f_X(x)} \\ & \cancel{+ \binom{n}{r+1}(r+1) f_X(x) F_X(x)^r (1-F_X(x))^{n-r-1}} \cancel{- (\quad)} \\ & \cancel{+ (\quad)} \cancel{- (\quad)} \\ & \cancel{+ (\quad)} - \,-\, - \cancel{- (\quad)} \\ & \cancel{+ (\quad)} \qquad \cancel{+ (n-n) * (\quad)} = 0 \end{aligned}$$

<u>Huge</u> cancellation happens!! and, because of the equality $\binom{n}{r}(n-r) = \binom{n}{r+1}(r+1)$ each of the summands after the first one disappears. Hence

$$f_{X_{(r)}}(x) = \frac{n!}{(r-1)!(n-r)!} f_X(x) F_X(x)^{r-1} (1 - F_X(x))^{n-r}$$ holds

4) I also discussed the idea of the proof of Theorem 7.4 on p. 62. Again, we first get the cdf and then find the mixed partial derivative

$$\frac{\partial^2}{\partial u \partial v} F_{X_{(i)},X_{(j)}}(u,v) \text{ to calculate the density } f_{X_{(i)},X_{(j)}}(u,v).$$

With the <u>discrete</u> variables U and V as introduced on p. 62 we see that

$$(U, V, n-U-V) \sim \text{Multinomial}(n;\ F_X(u),\ F_X(v)-F_X(u),\ 1-F_X(v))$$

Then we observe that

$$F_{X_{(i)},X_{(j)}}(u,v) = P(U \geq i \cap U+V \geq j) =$$

$$= \sum_{k=i}^{j-1} \sum_{m=j-k}^{n-k} P(U=k, V=m) + P(U \geq j)$$

Since the second summand does <u>not</u> involve v, its mixed partial derivative w.r. <u>u and v</u> will be <u>zero</u> and hence

$$f_{X_{(i)},X_{(j)}}(u,v) = \frac{\partial^2}{\partial u \partial v} \sum_{k=i}^{j-1} \sum_{m=j-k}^{n-k} \frac{n!}{k!\,m!\,(n-k-m)!} (F_X(u))^k (F_X(v)-F_X(u))^m (1-F_X(v))^{n-k-m}$$

Again, a huge cancellation happens when we calculate the partial derivatives by using the product rule for differentiation and we end up with

$$f_{X_{(i)},X_{(j)}}(u,v) = \frac{n!}{(i-1)!\,(j-1-i)!\,(n-j)!} f_X(u) f_X(v) (F_X(u))^{i-1} (F_X(v)-F_X(u))^{j-1-i} (1-F_X(v))^{n-j}$$

for $n \geq j > i \geq 1$, $\quad -\infty < u < v < \infty$ (and $= 0$ else)

(5) I also discussed in detail the example stating that for the range $R = X_{(n)} - X_{(1)}$ for order statistic $X_{(1)} < X_{(2)} < \dots < X_{(n)}$ from the uniform $(0,1)$ distribution, it holds $f_R(u) = \begin{cases} n(n-1)u^{n-2}(1-u), & 0<u<1 \\ 0 & \text{else} \end{cases}$

Proof:

To this end, first we note that by using the formula from Theorem 7.4 we have (with $i=1$ and $j=n$):

$$f_{X_{(1)},X_{(n)}}(x,y) = n(n-1)(F_X(y)-F_X(x))^{n-2} f_X(x) f_X(y)$$

We introduce the variable of interest

$U = X_{(n)} - X_{(1)}$ and one auxiliary variable

$V = X_{(n)}$ so that we could apply

the density transformation formula

$$f_{\binom{U}{V}}(u,v) = f_{\binom{X_{(1)}}{X_{(n)}}}(x(u,v), y(u,v)) \left| \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} \right|$$

Note: $\begin{vmatrix} X_{(1)} = V - U =: x \text{ and} \\ X_{(n)} = V =: y \end{vmatrix}$ $\left| \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} \right| = \left| \begin{vmatrix} -1 & 1 \\ 0 & 1 \end{vmatrix} \right| = 1$

Hence $f_{\binom{U}{V}}(u,v) = n(n-1)(F_X(v)-F_X(v-u))^{n-2} f_X(v-u) f_X(v) * 1$

To get $f_U(u)$ (which we are interested in) we need to integrate out the unwanted variable V from the joint density $f_{\binom{U}{V}}(u,v)$. We need to be careful with the integration range when doing this:

Since $0 < X_{(1)} < X_{(n)} < 1$ we get

$$0 < V - U < V < 1$$

$$\underline{0 < U < V < 1}$$

. This means that for a *fixed* U, V ranges in the interval $(u, 1)$.

Therefore:

$$f_R(u) = \int_u^1 f_{\binom{U}{V}}(u,v)\,dv = \int_u^1 n(n-1)\,(V-(V-u))^{n-2} * 1\,dv$$

$$= \begin{cases} n(n-1)u^{n-2}(1-u) & \text{if } 0 < u < 1 \\ 0 & \text{else} \end{cases}$$

I also advised you to repeat this exercise by using $X_{(1)} = V$ as an auxiliary variable. The intermediate calculations will be slightly different but at the end after integrating out V again (BUT THIS TIME in the range $(0, 1-u)$ (!)) you will get the same final result for the density $f_R(u)$.

6.) I also discussed one more problem in class (7/d) from tutorial sheet 4) but because it is completely solved in the solutions to tutorial set 4, I abstain from reproducing the derivation here.