Answers to the 9th problem set

1. It is desirable to make widgets produced on an assembly line as nearly identical as possible. If $X_1, \ldots, X_n$ are the logarithms of the masses n randomly chosen widgets, and $X_1, \ldots, X_n \sim$ i. i. d. $N(\mu, \sigma^2)$, then the above desideratum means making σ as small as possible.

   Recall that the normal density is

$$\varphi_{\mu,\sigma^2}(x) = \text{constant} \cdot \frac{1}{\sigma} \exp\left(\frac{-1}{2} \left(\frac{x-\mu}{\sigma} \right)^2 \right)$$

   and that

$$\sum_{i=1}^{n} (x_i - \mu)^2 = n(\overline{x} - \mu)^2 + \underbrace{\sum_{i=1}^{n} (x_i - \overline{x})^2}_{} = n\left[(\overline{x} - \mu)^2 + s^2 \right].$$

$$\uparrow \quad \boxed{\text{This defines } s^2.}$$

   (a) Suppose $0 < \sigma_0 < \sigma_1$.

   In order to test the null hypothesis $H_0 : \sigma = \sigma_0$
   against the alternative hypothesis $H_1 : \sigma = \sigma_1$

   we have used the likelihood-ratio statistic

$$\Lambda(x_1, \ldots, x_n) = \frac{f_{X_1,\ldots,X_n}(x_1, \ldots, x_n \mid \sigma = \sigma_1)}{f_{X_1,\ldots,X_n}(x_1, \ldots, x_n \mid \sigma = \sigma_0)}.$$

   Show that $\Lambda(x_1, \ldots, x_n)$ is an increasing function of s^2.

   <u>Answer</u>: I should probably have mentioned that "μ" should be replaced by its maximum-likelihood estimate, since it is unrealistic at best to assume μ is known with certainty. But fortunately, that does not change the bottom-line result.

$$\Lambda(x_1, \ldots, x_n) = \frac{f_{X_1,\ldots,X_n}(x_1, \ldots, x_n \mid \sigma = \sigma_1)}{f_{X_1,\ldots,X_n}(x_1, \ldots, x_n \mid \sigma = \sigma_0)} \overset{\boxed{\text{Because of independence}}}{=} \frac{\prod_{i=1}^{n} f_{X_i}(x_i \mid \sigma = \sigma_1)}{\prod_{i=1}^{n} f_{X_i}(x_i \mid \sigma = \sigma_0)}$$

$$= \prod_{i=1}^{n} \frac{f_{X_i}(x_i \mid \sigma = \sigma_1)}{f_{X_i}(x_i \mid \sigma = \sigma_0)} = \prod_{i=1}^{n} \frac{\dfrac{1}{\sigma_1} \exp\left(\dfrac{-1}{2} \left(\dfrac{x_i - \mu}{\sigma_1} \right)^2 \right)}{\dfrac{1}{\sigma_0} \exp\left(\dfrac{-1}{2} \left(\dfrac{x_i - \mu}{\sigma_0} \right)^2 \right)}$$

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

CONTINUED⟶

On the previous page we have shown that

$$\Lambda(x_1,\ldots,x_n) = \frac{\sigma_0^n}{\sigma_1^n}\exp\left(\frac{\sigma_1-\sigma_0}{2\sigma_1\sigma_0}\sum_{i=1}^n (x_i-\mu)^2\right).$$

Because $\dfrac{\sigma_0^n}{\sigma_1^n} > 0$, and exp is an increasing function, and $\dfrac{\sigma_1-\sigma_0}{2\sigma_1\sigma_0} > 0$, we can conclude that this expression is an increasing function of

$$\sum_{i=1}^n (x_i-\mu)^2 = n((\overline{x}-\mu)^2+s^2).$$

If we replace μ by its maximum-likelihood estimate $\widehat{\mu} = \overline{x}$, then the above is just $ns^2 = \sum_{i=1}^n (x_i-\overline{x})^2$, but regardless of whether we do that or not, it is an increasing function of s^2.

(b) Let H_0 and H_1 be as in part (a). Let capital $S =$ the same function of $X_1,\ldots,X_n$ that lower-case s is of $x_1,\ldots,x_n$. Let $K = K(X_1,\ldots,X_n)$ be some other statistic. Suppose $\mathbf{Pr}(K > \kappa \mid H_0) = 0.03$ and $\mathbf{Pr}(\chi^2_{n-1} > \ell) = 0.03$. How do you know that the test that rejects H_0 if and only if $nS^2 > \ell$ is at least as powerful as the one that rejects H_0 if and only if $K > \kappa$?

<u>**Answer:**</u> As noted in class somewhat before this problem set was due, this should have said:

"...the test that rejects H_0 if and only if $nS^2/\sigma_0^2 > \ell$ is at least as powerful..."

The likelihood ratio test is the test that rejects H_0 iff $\Lambda > \lambda_0$, where λ_0 is so chosen that $\mathbf{Pr}(\Lambda > \lambda_0 \mid H_0) = 0.03$. Since Λ is an increasing function of S^2, and so also an increasing function of nS^2/σ_0^2, there is some number ℓ such that $nS^2/\sigma_0^2 > \ell$ if and only if $\Lambda > \lambda_0$. Since $nS^2/\sigma_0^2 \sim \chi^2_{n-1}$ if H_0 is true, that number is the same as the number ℓ such that $\mathbf{Pr}(\chi^2_{n-1} > \ell) = 0.03$. Consequently the test that rejects H_0 if and only if nS^2/σ_0^2 is the likelihood ratio test. The Neyman-Pearson Lemma, proved in the answer to #12 on the 8th problem set, then implies that this test is more powerful than any other.

CONTINUED⟶

(c) Suppose $0 < \sigma_0 < \sigma_1$. Suppose $K = K(X_1, \ldots, X_n)$ is some statistic that is worthless for testing hypotheses about the value of σ because the probability distribution of K in no way depends on σ. Suppose $\mathbf{Pr}(K > \kappa) \le 0.03$. We indicate the lack of dependence on σ by saying

"$\alpha = \mathbf{Pr}_\sigma(K > \kappa)$ is the same regardless of the value of σ."

Choose m so that $\mathbf{Pr}(\chi^2_{n-1} > m) = \alpha$. Pretend you are completely ignorant of the nature of the chi-square distribution, you know nothing about the distribution of S^2 except that, as it says in parts (a) and (b) above, S^2 is an increasing function of Λ, and you otherwise know the results of parts (a) and (b) above, and of #12 on the 8th problem set. Explain how, in this state of ignorance, you would justify each step labelled with a "?" below.

$$P_{\sigma_0}(\underset{\uparrow}{nS^2/\sigma_0^2}> \ell) \overset{\substack{\boxed{?}\\ \downarrow}}{=} P_{\sigma_0}(K > \kappa) \overset{\substack{\boxed{?}\\ \downarrow}}{=} P_{\sigma_1}(K > \kappa) \overset{\substack{\boxed{?}\\ \downarrow}}{\le} P_{\sigma_1}(\underset{\uparrow}{nS^2/\sigma_0^2}> \ell).$$

As in part (b), this denominator should be here.

. . . and here.

(Summary: $P_\sigma(nS^2/\sigma_0^2 > \ell)$ is an increasing function of σ.)

Answer: Again, "$\ldots/\sigma_0^2\ldots$" was omitted; it is added below.

$P_{\sigma_0}(nS^2/\sigma_0^2 > \ell) = \alpha$ because $nS^2/\sigma_0^2 \sim \chi^2_{n-1}$ and $\mathbf{Pr}(\chi^2_{n-1} > \ell) = \alpha$.

That $P_{\sigma_0}(K > \kappa) = \alpha$ was given. Therefore the first equality holds.

The second equality follows from the fact that $\mathbf{Pr}(K > \kappa)$ does not depend on σ.

The third relation — an inequality — is the "hard part." It follows from the Neyman-Pearson Lemma, proved in #12 on the 8th problem set, in conjunction with our conclusion in part (b) that the test based on S^2 is the likelihood-ratio test.

(d) Suppose the null and alternative hypotheses are:

$$\begin{aligned} H_0 : \sigma &\leq 1, \\ H_1 : \sigma &> 1. \end{aligned}$$

Observe that these hypotheses make sense in our assembly-line scenario. Let K be some statistic such that $\mathbf{Pr}_\sigma(K > \kappa) \leq 0.03$ whenever $\sigma \leq 1$. Show that for any $\sigma_1 > 1$ we

"$\leq$", not "$=$".

have $\mathbf{Pr}_{\sigma_1}(nS^2 > \ell) \geq \mathbf{Pr}_{\sigma_1}(K > \kappa)$, i.e., the test based on S^2 is at least as powerful as the test based on K.

Answer: First note that we do not divide by anything called "σ_0^2" this time; our test statistic is just nS^2.

If $\mathbf{Pr}_\sigma(K > \kappa) \leq 0.03$ whenever $\sigma \leq 1$, then *a fortiori* $\mathbf{Pr}_\sigma(K > \kappa) \leq 0.03$ when $\sigma = 1$. The Neyman-Pearson Lemma then implies that if $\mathbf{Pr}_{\sigma=1}(K > \kappa) \leq 0.03$, then $\mathbf{Pr}_{\sigma_1}(nS^2 > \ell) \geq \mathbf{Pr}_{\sigma_1}(K > \kappa)$. Given the way the problem was stated, that is a sufficient answer.

But some difficulties occasioned by this question made me realize that in haste I did not write all of what I meant to write. What I intended was not just that $[nS^2 > \ell]$ is more powerful at $\sigma_1 > 1$ than any test whose power is ≤ 0.03 when $\sigma \leq 1$. Rather I had in mind that $[nS^2 > \ell]$ is itself one of those tests having power ≤ 0.03 whenever $\sigma \leq 1$, and therefore that it is the most powerful at $\sigma_1 > 1$ among all tests having power ≤ 0.03 whenever $\sigma \leq 1$. To do that, use the result of part (c). The result of part (c) implies that $\mathbf{Pr}_\sigma(nS^2/1^2 > \ell)$ gets bigger as σ gets bigger and smaller as σ gets smaller. Therefore, whenever $\sigma \leq 1$, then $\mathbf{Pr}_\sigma(nS^2/1^2 > \ell) \leq 0.03 = \mathbf{Pr}_{\sigma=1}(nS^2/1^2 > \ell)$.

2. DeGroot & Schervish, p. 541, #4.

Answer: The chi-square test statistic is $\sum \dfrac{(\text{observed} - \text{expected})^2}{\text{expected}}$. If the null hypothesis is true, then this has a chi-square distribution with two degrees of freedom. Since the hypothesized proportions are 1/4, 1/2, and 1/4, the "expected" counts are:

$$\begin{aligned} 24 \cdot (1/4) &= 6, \\ 24 \cdot (1/2) &= 12, \\ \text{and } 24 \cdot (1/4) &= 6. \end{aligned}$$

Therefore the value of the test statistic is

$$\frac{(10-6)^2}{6} + \frac{(10-12)^2}{12} + \frac{(4-6)^2}{6} = \frac{11}{3} = 3.6666\ldots.$$

According to the table on page 775 of DeGroot & Schervish, $\mathbf{Pr}(\chi_2^2 > 4.605) = 0.9$. Since $3.6666\ldots < 4.605$, we cannot reject the null hypothesis at the 10% level. Generally, allowing a probability of Type I error to be more than 10% would be reckless, so we do not reject the null hypothesis.

3. DeGroot & Schervish, p. 548, #2.

Answer: The null hypothesis does not specify a value of θ, so θ must be estimated. The null hypothesis says $X_1, \ldots, X_{200} \sim \text{i. i. d. Bin}(4, \theta)$. The sum $X_1 + \cdots + X_n$ is sufficient for θ, and its distribution is $\text{Bin}(800, \theta)$. The maximum-likelihood estimator of θ is therefore $(X_1 + \cdots + X_{200})/800$. The sum $X_1 + \cdots + X_{200}$ is the sum of 33 0s, 67 1s, 66 2s, 15 3s, and 19 4s, so it is $33 \cdot 0 + 67 \cdot 1 + 66 \cdot 2 + 15 \cdot 3 + 19 \cdot 4 = 329$, so we have $\widehat{\theta} = 329/800 = 0.41125$. The estimated expected counts in the five cells are therefore

$$\begin{aligned} 200(1-\widehat{\theta}\,)^4 = 200(1-329/800)^4 &\cong 24.029995 \\ 200 \cdot 4\widehat{\theta}(1-\widehat{\theta}\,)^3 = 200 \cdot 4(329/800)(1-329/800)^3 &\cong 67.14113 \\ 200 \cdot 6\widehat{\theta}^2(1-\widehat{\theta}\,)^2 = 200 \cdot 6(329/800)^2(1-329/800)^2 &\cong 70.34851 \\ 200 \cdot 4\hat{\theta}^3(1-\widehat{\theta}\,) = 200 \cdot 4(329/800)^3(1-329/800) &\cong 32.7596 \\ 200 \cdot \hat{\theta}^4 = 200 \cdot (329/800)^4 &\cong 5.7207588 \end{aligned}$$

and the test statistic is

$$\begin{aligned} \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}} &= \frac{(33-24.029995)^2}{24.029995} + \frac{(67-67.14113)^2}{67.14113} + \frac{(66-70.34851)^2}{70.34851} \\ &\quad + \frac{(15-32.7596)^2}{32.7596} + \frac{(19-5.7207588)^2}{5.7207588} \cong 44.7. \end{aligned}$$

Under the null hypothesis, the distribution of the test statistic should be χ^2_3.

NOTA BENE: THREE degrees of freedom. There are five categories; we lose one degree of freedom because the sum of the counts is fixed at 200 and not random, and another because θ was estimated based on the data.

According to the table on page 775, $\mathbf{Pr}(\chi^2_3 > 12.84) = 0.995$, and 44.7 is much bigger than 12.48, so we would reject the null hypothesis even if we tolerate only a very tiny probability of Type I error.

4. DeGroot & Schervish, p. 672, #5.

Answer: This book establishes (on page 397) a convention of taking "sample variance" to mean $(1/n)\sum_{i=1}^n (Y_i - \overline{Y})^2$ rather than $(1/(n-1))\sum_{i=1}^n (Y_i - \overline{Y})^2$. Therefore if we multiply the "sample variances" by 10, we will get sums of squares within groups, whose sum is the sum of squares due to "error," which is therefore $303 + 544 + 25 + 364 = 1236$.

source	d.f.	sum of squares	mean square	F
groups	3	$10\sum_{i=1}^4 \left(\overline{Y}_{i\bullet} - \overline{Y}_{\bullet\bullet}\right)^2 = 1593.8$	$531.2666\ldots$	$15.4737864\ldots$
error	36	$\sum_{i=1}^4 \sum_{j=1}^{10} \left(Y_{ij} - \overline{Y}_{i\bullet}\right)^2 = 1236.$	$34.333\ldots$	

According to the table on page 780, the 97.5^{th} percentile of the $F_{3,30}$ distribution is 3.59, so the same percentile of the $F_{3,36}$ distribution must be even smaller, and thus, far, far smaller than the observed value, $15.47\ldots$. The null hypothesis of equality of the four average scores is therefore overwhelmingly rejected.

The theory on which we have based all of this relies on the assumptions that the four population variances are equal, that the observations are independent (that would result from the way in which the samples were chosen, not from any presumed independence of one student's score and another's), and that these are normally distributed populations.

5. Recall that for $0 < p < 1$ we have $\operatorname{logit}(p) = \log \dfrac{p}{1-p}$. Suppose $X \mid [\mu = \mu_i] \sim N(\mu_i, 1^2)$ for $i = 1, 2$. A prior probability distribution is assigned to μ, so that $\mathbf{Pr}(\mu = \mu_1) + \mathbf{Pr}(\mu = \mu_2) = 1$. Show that for some A, B,

$$\operatorname{logit} \mathbf{Pr}(\mu = \mu_1 \mid X = x) = Ax + B + \operatorname{logit} \mathbf{Pr}(\mu = \mu_1),$$

i.e., the logit of the posterior probability is some function of x whose graph is a straight line plus the logit of the prior probability. Find the values of A and B.

Answer: According to Bayes' formula, we multiply the likelihood by the prior and then normalize, to get the posterior:

$$\begin{aligned}
&\underbrace{(\mathbf{Pr}(\mu = \mu_1 \mid X = x), \mathbf{Pr}(\mu = \mu_2 \mid X = x))}_{\text{posterior}} \\
= \ &[\text{constant}] \cdot \underbrace{(\mathbf{Pr}(\mu = \mu_1), \mathbf{Pr}(\mu = \mu_2))}_{\text{prior}} \cdot \underbrace{(f_{X|\mu=\mu_1}(x), f_{X|\mu=\mu_2}(x))}_{\text{likelihood}}
\end{aligned}$$

("constant" in this case means not depending on whether $i = 1$ or $i = 2$.)

So $\quad \operatorname{logit} \mathbf{Pr}(\mu = \mu_1 \mid X = x)$

$$= \log \frac{\mathbf{Pr}(\mu = \mu_1 \mid X = x)}{1 - \mathbf{Pr}(\mu = \mu_1 \mid X = x)} = \log \frac{\mathbf{Pr}(\mu = \mu_1 \mid X = x)}{\mathbf{Pr}(\mu = \mu_2 \mid X = x)}$$

$$= \underbrace{\log \left(\frac{[\text{constant}] \cdot \mathbf{Pr}(\mu = \mu_1) f_{X|\mu=\mu_1}(x)}{[(\text{same})\ \text{constant}] \cdot \mathbf{Pr}(\mu = \mu_2) f_{X|\mu=\mu_2}(x)} \right)} = \log \frac{\mathbf{Pr}(\mu = \mu_1)}{\mathbf{Pr}(\mu = \mu_2)} + \log \frac{f_{X|\mu=\mu_1}(x)}{f_{X|\mu=\mu_2}(x)}$$

The word "same" in this case is redundant. The "constant" is "constant" only because it's the same in both cases.

$$= \log \frac{\mathbf{Pr}(\mu = \mu_1)}{1 - \mathbf{Pr}(\mu = \mu_1)} + \log \frac{f_{X|\mu=\mu_1}(x)}{f_{X|\mu=\mu_2}(x)} = \operatorname{logit} \mathbf{Pr}(\mu = \mu_1) + \log \frac{f_{X|\mu=\mu_1}(x)}{f_{X|\mu=\mu_2}(x)}$$

So we need to show that

$$\log \frac{f_{X|\mu=\mu_1}(x)}{f_{X|\mu=\mu_2}(x)} = Ax + B$$

for some A and B, and find the values of A and B. We have

$$\log \frac{f_{X|\mu=\mu_1}(x)}{f_{X|\mu=\mu_2}(x)} = \log \frac{\dfrac{1}{\sqrt{2\pi}} \exp\left(\dfrac{-(x-\mu_1)^2}{2}\right)}{\dfrac{1}{\sqrt{2\pi}} \exp\left(\dfrac{-(x-\mu_2)^2}{2}\right)} = \log \exp\left(\left(\frac{-(x-\mu_1)^2}{2}\right) - \left(\frac{-(x-\mu_2)^2}{2}\right) \right)$$

$$= \frac{-(x^2 - 2x\mu_1 + \mu_1^2)}{2} - \frac{-(x^2 - 2x\mu_2 + \mu_2^2)}{2} = (\mu_1 - \mu_2)x - (\mu_1^2 - \mu_2^2).$$

So $A = \mu_1 - \mu_2$ and $B = \mu_1^2 - \mu_2^2$. It may be worth noticing that this is

$$(\mu_1 - \mu_2)x - (\mu_1^2 - \mu_2^2) = (\mu_1 - \mu_2)(x - \overline{\mu})$$

where $\overline{\mu} = (\mu_1 + \mu_2)/2$, so that we have a center at $\overline{\mu}$; if x is observed to be at the center, then the posterior probability is no different from the prior probability.