

Solutions to Practice Final
from March 2004

1. The observed values are as given. Under the null hypothesis of a fair die, the expected values are all $100/6 \approx 16.7$. The Chi-squared statistic can be computed as

$$(6 - 16.7)^2/16.7 + (12 - 16.7)^2/16.7 + \dots \approx 31.8 .$$

The appropriate degrees of freedom is 5. At level .05, you would clearly reject.

For (b) and (c), we need to compute the expected values under the hypothesis that opposite faces have the same probabilities. Under this hypothesis, let π_1 be the probability of observing a 1 or 6, π_2 the probability of observing a 2 or 5, and π_3 a 3 or 4. The MLE for π_1 is the proportion in categories 1 or 6 or $(6 + 14)/100 = .2$, so the estimated expected number of outcomes that are 1 or 6 would be 20. The estimated expected number of outcomes that are 1 would be 10, and the same for 6. Similarly, the expected number of outcomes that are 2 or 5 is .3. So the estimated expected table would be (from 1-6): 10, 15, 25, 25, 15, 10. The Chi-squared statistic becomes

$$(6 - 10)^2/10 + (12 - 15)^2/15 + \dots = 6.4 .$$

The degrees of freedom is $(6 - 1) - (3 - 1) = 3$. At level .05, the critical value is 9.35, and so the null is not rejected.

3. Under $\theta = 1$,

$$P_{\theta=1}\{\bar{X}_n \geq c_n\} = P_{\theta=1}\{n^{1/2}(\bar{X}_n - 1) \geq n^{1/2}(c_n - 1)\} .$$

Using the normal approximation, it follows that

$$n^{1/2}(c_n - 1) \approx z_{1-\alpha} = 2.33 .$$

Solving for $c_n = 1 + 2.33/n^{1/2}$.

Under $\theta = 1.21$, the probability of a type 2 error is

$$P_{\theta=1.21}\{\bar{X}_n < c_n\} = P_{\theta=1.21}\left\{n^{1/2}\frac{(\bar{X}_n - 1.21)}{1.1} < n^{1/2}(c_n - 1.21)/1.1\right\} .$$

In order for this to be 5 percent, set the right side equal to the 5th percentile of $N(0, 1)$, which is -1.645. So solve

$$n^{1/2}(c_n - 1.21)/1.1 = -1.645$$

for n . Plug in the previous expression for c_n here, and solving for n gives about 468.

3. You are not responsible for this problem, since we did neither the sign test nor the wilcoxon signed rank test for paired data.

4. By simple integration, $E(X_i) = \theta + 1$. Set this equal to \bar{X}_n and you get $\hat{\theta}_{MM} = \bar{X}_n - 1$. This estimator is unbiased for θ , and so its mean squared error is just $Var(\bar{X}_n) = Var(X_i)/n = 1/n$.

The likelihood function is

$$\prod_i \exp[-(X_i - \theta)] I\{X_i \geq \theta\} .$$

Let $X_{(1)}$ denote the minimum of (X_1, \dots, X_n) . Then the product of the indicator terms in the likelihood is 1 iff $X_{(1)} \geq \theta$. The likelihood becomes

$$\exp(-\sum_i X_i - n\theta) \cdot I\{X_{(1)} \geq \theta\} .$$

Graph this function to get the MLE is $\hat{\theta}_n = X_{(1)}$. (Simply observe that the likelihood without the indicator event is an increasing function of θ , but θ is constrained to be no bigger than $X_{(1)}$ or the likelihood defaults to zero.)

For $t > \theta$, the cdf of X_i is $1 - \exp[-(t - \theta)]$.

$$P_\theta\{X_{(1)} < t\} = 1 - P_\theta^n(X_i > t) = 1 - \exp[-n(t - \theta)] .$$

From here, differentiate to compute the density of the MLE, and then its mean and variance to compute its MSE.

5. This is just a Chi-squared test of homogeneity in two by two contingency table. The observed entries are 100, 50, 900, 950, with corresponding “estimated” entries 75, 75, 925, 925. The Chi-squared statistic is 18.05, and with one degree of freedom is significant.

6. Note there is a typo in the problem. The σ_2 should be σ^2 . Here there are $n + 1$ parameters: μ_1, \dots, μ_n and σ . The likelihood function is

$$\prod_i (2\pi\sigma^2)^{-1/2} \exp[-(X_i - \mu_i)^2/(2\sigma^2)] (2\pi\sigma^2)^{-1/2} \exp[-(Y_i - \mu_i)^2/(2\sigma^2)]$$

Taking logs and partial derivatives with respect to each of the parameters leads to the MLES $\hat{\mu}_i = (X_i + Y_i)/2$ and

$$\hat{\sigma}^2 = \frac{1}{4n} \sum_i (X_i - Y_i)^2 .$$

By linearity,

$$E(\hat{\sigma}^2) = \frac{1}{4} \sum_i E[(X_i - Y_i)^2] = \frac{1}{4} \sum_i \text{Var}(X_i - Y_i) ,$$

since $X_i - Y_i$ has mean 0. But, $\text{Var}(X_i - Y_i) = \text{Var}(X_i) + \text{Var}(Y_i) = 2\sigma^2$, and so

$$E(\hat{\sigma}^2) = \sigma^2/2 .$$

Finally, the estimator is an average of i.i.d. random variables $Z_i = \frac{1}{4}(X_i - Y_i)^2$. By the law of large numbers $\hat{\sigma}^2 = \bar{Z}_n$ converges in probability to $E(Z_i) = E(X_i - Y_i)^2/4$, which is $\sigma^2/2$. Since the limit is not σ^2 , the estimator is inconsistent.