STATS 2107
Statistical Modelling and Inference II
Tutorial 1
Solutions

Jono Tuke, Matt Ryan

Semester 2 2022

1. Suppose $Y_1, Y_2, \ldots, Y_n$ are i.i.d. $N(\mu, \sigma^2)$ random variables and let $\bar{Y}$ denote the sample mean. Using moment generating functions (MGFs), prove that $\bar{Y} \sim N(\mu, \sigma^2/n)$.

Solutions:
First consider the moment generating function for a general sample mean:

$$
\begin{aligned}
M_{\bar{Y}}(t) &= E[e^{t\bar{Y}}] \\
&= E[e^{t(Y_1/n + Y_2/n + \ldots + Y_n/n)}] \\
&= E[e^{tY_1/n + tY_2/n + \ldots + tY_n/n}] \\
&= E[e^{tY_1/n}]E[e^{tY_2/n}]\ldots E[e^{tY_n/n}] \qquad \text{by independence} \\
&= M_{Y_1}(t/n)M_{Y_2}(t/n)\ldots M_{Y_n}(t/n) \\
&= \prod_{i=1}^{n} M_{Y_i}(t/n). \qquad (*)
\end{aligned}
$$

We are given that $Y_i \sim N(\mu, \sigma^2)$ which has MGF

$$
M_{Y_i}(t) = \exp\left\{\mu t + \frac{t^2\sigma^2}{2}\right\}
$$

Using (*) we have

$$
\begin{aligned}
M_{\bar{Y}}(t) &= \left[\exp\left\{\frac{\mu t}{n} + \frac{t^2\sigma^2}{2n^2}\right\}\right]^n \\
&= \exp\left\{\sum_{i=1}^{n}\frac{\mu t}{n} + \sum_{i=1}^{n}\frac{t^2\sigma^2}{2n^2}\right\} \\
&= \exp\left\{\mu t + \frac{t^2\sigma^2}{2n}\right\}.
\end{aligned}
$$

Which by the uniqueness of MGFs proves that $\bar{Y}$ is normally distributed with mean μ and variance σ^2/n.

---

2. Let $Z_1, Z_2, \ldots, Z_p$ be i.i.d. $N(0,1)$ random variables.

a. Show that the moment generating function $M_{Z_i^2}(t)$ of Z_i^2 is given by $(1-2t)^{-\frac{1}{2}}$ for each $i = 1, 2, \ldots, p$.

Solutions:

$$\begin{aligned}
M_{Z_i^2}(t) &= E[e^{tZ_i^2}] \\
&= \int_{-\infty}^{\infty} e^{tz^2} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{z^2(t-1/2)} dz \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2(1-2t)/2} dz \\
&= \frac{1}{(1-2t)^{1/2}} \int_{-\infty}^{\infty} \frac{(1-2t)^{1/2}}{\sqrt{2\pi}} e^{-z^2(1-2t)/2} dz
\end{aligned}$$

but the integral is one since it is the pdf of $N(0, \frac{1}{1-2t})$. So we have

$$M_{Z^2}(t) = \frac{1}{(1-2t)^{1/2}}.$$

---

b. Hence, show that the moment generating function of $X = \sum_{i=1}^{p} Z_i^2$ is

$$M_X(t) = \frac{1}{(1-2t)^{p/2}}.$$

Solutions:
Using independence of the Z_i we can conclude that the Z_i^2 are independent. Then using the property of independent MGFs we get that:

$$\begin{aligned}
M_X(t) &= M_{\sum_{i=1}^{p} Z_i^2}(t) \\
&= \prod_{i=1}^{p} M_{Z_i^2}(t) \\
&= \prod_{i=1}^{p} (1-2t)^{-\frac{1}{2}} \\
&= \frac{1}{(1-2t)^{p/2}}.
\end{aligned}$$

---

3. Let $Y \sim Bin(n, p)$ and consider two estimators for p; namely

$$\hat{p}_1 = \frac{Y}{n} \text{ and } \hat{p}_2 = \frac{Y+1}{n+2}.$$

a. Show that $\hat{p}_1$ is unbiased for p.

Solutions:

$$\begin{aligned} E[\hat{p}_1] &= E[Y/n] \\ &= \frac{1}{n}E[Y] \\ &= \frac{np}{n} \\ &= p. \end{aligned}$$

b. Derive the bias of $\hat{p}_2$.

Solutions:

$$\begin{aligned} b_{\hat{p}_2}(p) &= E[\hat{p}_2] - p \\ &= E\left[\frac{Y+1}{n+2}\right] - p \\ &= \frac{np+1}{n+2} - p \\ &= \frac{np+1-p(n+2)}{n+2} \\ &= \frac{1-2p}{n+2}. \end{aligned}$$

c. Find $MSE_{\hat{p}_1}(p)$ and $MSE_{\hat{p}_2}(p)$.

Solutions:

$$\begin{aligned} MSE_{\hat{p}_1}(p) &= \text{Var}(\hat{p}_1) + b_{\hat{p}_1}(p)^2 \\ &= \text{Var}(Y/n) + 0 \\ &= \frac{1}{n^2}\text{Var}(Y) \\ &= \frac{np(1-p)}{n^2} \\ &= \frac{p(1-p)}{n}. \\ MSE_{\hat{p}_2}(p) &= \text{Var}\left(\hat{p}_2\right) + b_{\hat{p}_2}(p)^2 \\ &= \text{Var}\left(\frac{Y+1}{n+2}\right) + \left(\frac{1-2p}{n+2}\right)^2 \\ &= \frac{\text{Var}(Y) + (1-2p)^2}{(n+2)^2} \\ &= \frac{np(1-p) + (1-2p)^2}{(n+2)^2}. \end{aligned}$$

d. If $p = 0.5$, which estimator has the largest MSE?

Solutions:

$$
\begin{aligned}
MSE_{\hat{p}_1}(p) &= \frac{1}{4n} \\
MSE_{\hat{p}_2}(p) &= \frac{n}{4(n+2)^2} \\
&= \frac{1}{4(n+4+4/n)}
\end{aligned}
$$

So $MSE_{\hat{p}_2}(p) < MSE_{\hat{p}_1}(p)$

```r
mse1  <- function(p,n){
  p*(1-p)/n
}
mse2  <- function(p,n){
  (n*p*(1-p) + (1-2*p)^2)/(n+2)^2
}
df  <- tibble(n = 1:100)
df  <- df %>% mutate(estimator1 = mse1(n, p = 0.5),
                     estimator2 = mse2(n, p = 0.5))
df %>% gather(key = "MSE", value = value, -n) %>%
  ggplot(aes(n, value, col = MSE)) + geom_line() +
  scale_color_brewer(palette = "Set1")
```