Maximum Likelihood Estimation and Model Fitting, CMEE MSc. Tin-Yu Hui

Practical 1 (22 Feb 2021)

Question 1

i. Let X be a Bernoulli random variable with success probability p. Show that $E(X) = p$ and $E(X^2) = p$.

Same as in lecture notes

ii. Calculate the variance of X from $E(X)$ and $E(X^2)$.

Question 2

i. Given a r.v. X with pdf $f_X(x) = \frac{2}{x^3}, 1 \le x < \infty$, show that $\int_{-\infty}^{\infty} f_X(x)dx = 1$.

$\int_{-\infty}^{+\infty} f(x)dx = \int_1^{+\infty} \frac{2}{x^3}dx$ (Note: the lower bound of the r.v. support)

$= \int_1^{\infty} 2x^{-3}dx = \frac{2x^{-2}}{-2}\Big|_1^{\infty} = -x^{-2}\Big|_1^{\infty}$

$= (-\infty^{-2}) - (-1^{-2})$

$= \left(-\frac{1}{\infty^2}\right) - (-1)$

$= 1$

Area under pdf = 1 means it is a proper pdf

Note: Normally we do not directly substitute ∞ into the equations.

ii. Please find $E(X)$.

$E(X) = \int_1^{\infty} x f(x)\,dx$

$= \int_1^{\infty} x(2x^{-3})dx$

$= \int_1^{\infty} 2x^{-2}dx$

$= \frac{2x^{-1}}{-1}\Big|_1^{\infty}$

$= \frac{-2}{x}\Big|_1^{\infty}$

$= \left(\frac{-2}{\infty}\right) - \left(\frac{-2}{1}\right)$

$= 0 - (-2)$

$= 2$

Question 3

i. Let $X \sim Poisson(\lambda)$. Calculate (by hand) its expected value $E(X)$

$$E(X) = \sum_{x=0}^{\infty} x\, f(x) = \sum_{x=0}^{\infty} x \cdot \frac{\lambda^x e^{-\lambda}}{x!} = \sum_{x=1}^{\infty} x \cdot \frac{\lambda^x e^{-\lambda}}{x!}$$

(∵ recognised that the first term on the summation is 0.

(∵ $\lambda^x = \lambda \cdot \lambda^{x-1}$, $\frac{x}{x!} = \frac{1}{(x-1)!}$)

$$= \sum_{x=1}^{\infty} \frac{\lambda \cdot \lambda^{x-1} e^{-\lambda}}{(x-1)!}$$

$$= \lambda \sum_{x=0}^{\infty} \frac{\lambda^{x-1} e^{-\lambda}}{(x-1)!}$$

ii. Let $X \sim Poisson(\lambda = 2)$. Remember that Poisson distributions model the number of events that had occurred within a time interval with an average rate of occurrence λ. Let us visualise its pmf in R using the `dpois()` function for x between 0 and 10:

```
x<-0:10
y<-dpois(x, lambda=2)
plot(x, y, pch=16, ylab='pmf', xlab='outcome')
```

Note that the Poisson pmf is still defined for $x > 10$, but the associated probabilities are too small that the pmf becomes less interesting.

$$= \lambda \sum_{y=0}^{\infty} \frac{\lambda^y e^{-\lambda}}{y!}$$ (sub $y = x-1$)

$$= \lambda \cdot 1$$ (Poisson pmf sum to one)

$$= \lambda$$

iii. Calculate $Pr(X = 4)$ and $Pr(X \leq 3)$, and verify your answers with R:

```
# Pr(X=4)
dpois(4, lambda=2)
# Pr(x<=3)
sum(dpois(0:3, lambda=2)
# OR EQUIVALENTLY
ppois(3, lambda=2)
```

iv. We can generate (simulate) Poisson random numbers using `rpois()`. Let us sample 1000 independent Poisson numbers with $\lambda = 2$:

```
x<-rpois(1000, lambda=2)
x
hist(x) # PLOT A HISTOGRAM
```

A different set of numbers will be generated when you rerun the `rpois()` command because of its random nature. In some circumstances (e.g. for debugging) wer would like to generate a "fixed" set of random numbers, which can be achieved by setting a random seed:

```
set.seed(123)
x<-rpois(10, lambda=2)
y<-rpois(10, lambda=2)
x==y # NOT ALL EQUAL
x
```

```
# RESET RANDOM SEED
set.seed(123)
z<-rpois(10, lambda=2)
x==z # SHOULD BE ALL EQUAL
z
```

Question 4

i. Suppose X is exponentially distributed with rate $\lambda > 0$. Clearly X is a continuous r.v.. The pdf is X is $f_X(x) = \lambda e^{-\lambda x}$, and its support is $[0, \infty)$. Show that $f_X(x)$ is a valid pdf (i.e. area under pdf=1).

$\int_0^\infty f_X(x)dx = \int_0^\infty \lambda e^{-\lambda x} dx = \left(-e^{-\lambda x}\right)\Big|_0^\infty$

$= \left(-e^{-\lambda(\infty)}\right) - \left(-e^{-\lambda(0)}\right)$

$= -0 - (-1) = 1$

ii. Calculate $E(X)$. (Integration by parts) Let $u = x$, $V = -e^{-\lambda x}$; $u' = 1$, $V' = \lambda e^{-\lambda x}$

$E(X) = \int_0^\infty x f(x)\, dx = \int_0^\infty x \lambda e^{-\lambda x} dx = \int_0^\infty u v' dx$

$= uv\Big|_0^\infty - \int_0^\infty u' v\, dx = \left(-xe^{-\lambda x}\Big|_0^\infty\right) + \int_0^\infty e^{-\lambda x} dx$

$= 0 + \int_0^\infty e^{-\lambda x} dx$

$= \frac{1}{\lambda}\int_0^\infty \lambda e^{-\lambda x} dx = \lambda$

iii. Show that the cumulative density function (cdf) of X is $F_X(x) = 1 - e^{-\lambda x}$.

$F_X(x) = \int_0^x f(t)\, dt = \int_0^x \lambda e^{-\lambda t} dt$ (t is a dummy variable)

$= -e^{-\lambda t}\Big|_0^x$

$= -e^{-\lambda x} - \left(-e^{-\lambda(0)}\right) = 1 - e^{-\lambda x}$

iv. Let $X \sim Exponential(\lambda = 2)$. Plot the pdf of X in R for $0 \le x \le 5$ with interval 0.01. These functions may be useful: `seq()`, `dexp()`, `plot()`

$-xe^{-\lambda x}\Big|_0^\infty = -\lim_{x\to\infty}\left(xe^{-\lambda x}\right) - \left(-0 \cdot e^{-\lambda(0)}\right) = -\lim_{x\to\infty} xe^{-\lambda x}$.

The problem is that, when $x \to \infty$, $e^{-\lambda x} \to 0$, ~~the product of the two terms~~ We need to decide which direction $(xe^{-\lambda x})$ will go to. Thankfully, we know that $e^{-\lambda x}$ goes to zero in a much faster rate (A LOT faster than $x \to \infty$ linearly). Therefore $(xe^{-\lambda x}) \to 0$ (See L'Hospital rule).

v. Calculating probabilities from pdfs. If X is a continuous r.v., then it is pointtless to calculate the probability of X taking on any certain value, as it is always zero. This is because there are infinitely many outcomes for a continuous r.v. and that you can never exactly hit a particular number. Instead, we can calculate the probability that X falls within an interval, say, $\Pr(a \leq X \leq b)$. And such probability is the area under the pdf between the interval: $\Pr(a \leq X \leq b) = \int_a^b f_X(x)dx$. Let us calculate the probability of $\Pr(0 \leq X \leq 1)$ for $X \sim Exp(\lambda = 2)$. Use `integrate()` for numerical integration.

```
integrate(dexp, lower=0, upper=1, rate=2)
```

Question 5

i. Let $X \sim N(\mu, \sigma^2)$. We call X the "standard normal" when $\mu = 0$ and $\sigma^2 = 1$. Plot the pdf of the standard normal distribution from $x = -3$ to $x = 3$.

ii. What is $\Pr(2 \leq X \leq 3)$? And what is $\Pr(-1.96 \leq X \leq 1.96)$? Use `integrate()` to help you with this.

iii. Verify your answer with the following commands:

```
pnorm(3)-prnorm(2)
pnorm(1.96)-pnorm(-1.96)
```

For the examples above it is safe to say that R knows a lot of distributions. For each distribution, we use the command with prefix `r` (e.g. `rnorm()`, `rpois()`, `rbinom()`) to generate random samples, and use the command with prefix `d` to calculate pmf/pdf. Those with prefix `p` (e.g. `pnorm()`) return the cdf of the distribution, and `q` for quantiles.

Question 6 [Central Limit Theorem, adopted from Mick Crawley's GLM course]

The central limit theorem (CLT) states that for any distribution with finite expected value and variance, the sample mean of the random samples from that distribution tends to be normally distributed.

Let us consider the negative binomial distribution. Generate 1000 negative binomial random numbers with $r = 1$ and $p = 0.2$ and plot them out.

```
y<-rnbinom(1000, 1, 0.2)

hist(y)
```

It is far from "normal" as it skews to a side. The CLT says the mean of samples will follow a normal distribution even for a badly behaved distribution like this. To visualise this effect let us consider the following codes:

```
# GENERATE 30000 NETAGIVE BINOMIAL RANDOM NUMBERS

# AND PUT THEM INTO A 1000-BY-30 MATRIX

y<-matrix(rnbinom(30*1000, 1, 0.2), nr=1000, nc=30)


# CALCULATE ROW MEAN

y.row.mean<-apply(y, 1, mean)


# PLOT THE HISTOGRAM OF THESE 1000 ROW MEANS

hist(y.row.mean)
```

Does the histogram look more "normal" now? This is why the normal distribution is so famous and widely used. The CLT holds for larger sample sizes, say, 30 or above.

Question 7 [Moment-generating function]

Let $X \sim N(\mu, \sigma^2)$. Find $E(X)$ and $E(X^2)$ from its mgf, where $M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$.

Note that $M_X(0) = e^{\mu(0) + \frac{1}{2}\sigma^2(0)} = e^{0+0} = 1$

t is a dummy variable here

$$M_X'(t) = \frac{d}{dt} M_X(t) = \frac{d}{dt} e^{\mu t + \frac{1}{2}\sigma^2 t^2} \quad (\text{differentiate w.r.t. } t)$$

$$= \left(e^{\mu t + \frac{1}{2}\sigma^2 t^2}\right) \frac{d}{dt}\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$$

$$= M_X(t)\left(\mu + \frac{1}{2}\sigma^2(2t)\right)$$

$$= M_X(t)(\mu + \sigma^2 t) \quad (*)$$

$$M_X'(0) = M_X(0)(\mu + \sigma^2(t)) \quad (\text{Sub } t=0)$$

$$= 1 \cdot (\mu + 0) = \mu$$

$\therefore E(X) = M_X'(0) = \mu$.

$$M_X''(t) = \frac{d}{dt} M_X'(t) = \frac{d}{dt}\left[M_X(t)(\mu + \sigma^2 t)\right]$$

$$= M_X(t) \frac{d}{dt}(\mu + \sigma^2 t) + (\mu + \sigma^2 t) M_X'(t) \quad (\because \text{Product rule})$$

$$= M_X(t)(0 + \sigma^2) + (\mu + \sigma^2 t) M_X'(t)$$

$$E(X^2) = M_X''(0) = M_X(0)(\sigma^2) + (\mu + \sigma^2(0)) M_X'(0)$$

$$= 1 \cdot \sigma^2 + \mu \cdot \mu = \mu^2 + \sigma^2$$

$\therefore E(X^2) = \mu^2 + \sigma^2$.

$$Var(X) = E(X^2) - (E(X))^2 = (\mu^2 + \sigma^2) - \mu^2 = \sigma^2$$

Extra: In fact, $M_X(t) = E(e^{Xt})$, the transformed ~~variable~~ r.v.