BST 140.651 Problem Set 2

Kai Aragaki

9/15/2020

Problem 1

$Var(X) = E[X^2] - E[X]^2$
Where $Var(X) = E[(X - \mu)^2]$, so
$E[(X - \mu)^2] = E[X^2] - E[X]^2$
$E[X^2 - 2X\mu + \mu^2] = E[X^2] - E[X]^2$
$E[X^2] + E[-2X\mu] + E[\mu^2] = E[X^2] - E[X]^2$
Since μ is a constant, and constants can be pulled out of the expected value function,
$E[X^2] - 2\mu E[X] + \mu^2 E[1] = E[X^2] - E[X]^2$
Simplifying, and with the knowledge that $E[X] = \mu$:
$-2\mu^2 + \mu^2 = -\mu^2$
From which we find
$-2\mu^2 = -2\mu^2$
Which we know to be true.

Problem 2

---

a

---

g is a valid density if all values are positive and if the integral across all values is 1.

We know all values are positive because it is the sum of three densities (which themselves must always be positive), each of which is multiplied by a scalar that is also positive.

To show that the integral across all values is equal to 1,
$\int_{-\infty}^{\infty} g(x) = 1$
Since we know
$g(x) = \pi_1 f_1(x) + \pi_2 f_2(x) + \pi_3 f_3(x)$
then
$\int_{-\infty}^{\infty} g(x) = \int_{-\infty}^{\infty} \left(\pi_1 f_1(x) + \pi_2 f_2(x) + \pi_3 f_3(x) \right)$
which is equivalent to
$\int_{-\infty}^{\infty} g(x) = \pi_1 \int_{-\infty}^{\infty} f_1(x) + \pi_2 \int_{-\infty}^{\infty} f_2(x) + \pi_3 \int_{-\infty}^{\infty} f_3(x)$

and since each function f_i is a valid density their integrals must equal 1, giving

$\int\limits_{-\infty}^{\infty} g(x) = \pi_1 + \pi_2 + \pi_3$

Since we are additionally told $\Sigma_{i=1}^{3}\pi_i = 1$

$\int\limits_{-\infty}^{\infty} g(x) = 1$ ∎

b

_____

We know that mean (or expected value) for a continous distribution is given by

$\int\limits_{-\infty}^{\infty} xg(x)$

Which is

$\int\limits_{-\infty}^{\infty} x\Big(\pi_1 f_1(x) + \pi_2 f_2(x) + \pi_3 f_3(x)\Big)$

which can be expanded to

$\pi_1 \int\limits_{-\infty}^{\infty} xf_1(x) + \pi_2 \int\limits_{-\infty}^{\infty} xf_2(x) + \pi_3 \int\limits_{-\infty}^{\infty} xf_3(x)$

And since we know the definition of the expected value, we can rewrite this as

$E[X] = \pi_1\mu_1 + \pi_2\mu_2 + \pi_3\mu_3$

c

_____

We have previously shown that the definition of

$Var(x) = E[X^2] - E[X]^2$

We can represent the variance of g in these terms:

$\int\limits_{-\infty}^{\infty} x^2 g(x) - [\int\limits_{-\infty}^{\infty} xg(x)]^2$

The right term we have previously found - it's simply the square of the expectation value we found in the previous problem. The left terms can be found by first writing g in terms of f

$\int\limits_{-\infty}^{\infty} x^2 g(x) = \int\limits_{-\infty}^{\infty} x^2\Big(\pi_1 f_1(x) + \pi_2 f_2(x) + \pi_3 f_3(x)\Big)$

or distributed:

$\int\limits_{-\infty}^{\infty} x^2 g(x) = \pi_1 \int\limits_{-\infty}^{\infty} x^2 f_1(x) + \pi_2 \int\limits_{-\infty}^{\infty} x^2 f_2(x) + \pi_3 \int\limits_{-\infty}^{\infty} x^2 f_3(x)$

and then noticing that the variance equation can be written as follows:

$\sigma^2 = E[X^2] + \mu^2$

And solving for $E[X^2]$:

$E[X^2] = \mu^2 + \sigma^2$

which is equivalent to

$$\int\limits_{-\infty}^{\infty} x^2 f_i(x) = \mu_i^2 + \sigma_i^2$$

Substituting in, we get

$$\int\limits_{-\infty}^{\infty} x^2 g(x) = \pi_1(\mu_1^2 + \sigma_1^2) + \pi_2(\mu_2^2 + \sigma_2^2) + \pi_3(\mu_3^2 + \sigma_3^2)$$

Combining this left hand term with the right hand term, we get

$$Var(x) = \pi_1(\mu_1^2 + \sigma_1^2) + \pi_2(\mu_2^2 + \sigma_2^2) + \pi_3(\mu_3^2 + \sigma_3^2) - (\pi_1\mu_1 + \pi_2\mu_2 + \pi_3\mu_3)^2$$

Problem 3

_____

a

_____

The mean of a continuous distribution is

$$\int\limits_{lower}^{upper} x f(x)$$

In this case the lower and upper bounds are given as 0 and 1 respectively.

Therefore the mean is given by

$$\int\limits_{0}^{1} (k+1)x^{k+1}$$

$$\int\limits_{0}^{1} kx^{k+1} + \int\limits_{0}^{1} x^{k+1}$$

$$\left.\frac{kx^{k+2}}{k+2}\right|_0^1 + \left.\frac{x^{k+2}}{k+2}\right|_0^1$$

$$\frac{k}{k+2} + \frac{1}{k+2}$$

$$\frac{k+1}{k+2}$$

b

_____

The variance of this distribution is

$$Var(x) = E[X^2] - E[X]^2$$

We know $E[X] = \mu$, which is what we found in part a, so squaring it is trivial. Additionally,

$$E[X^2] = \int\limits_{0}^{1} x^2\big((k+1)x^k\big) = \int\limits_{0}^{1} (k+1)x^{k+2}$$

And by distributing,

$$\int\limits_{0}^{1} kx^{k+2} + \int\limits_{0}^{1} x^{k+2}$$

We can see by analogy to the first problem that this will integrate and simplify to

$$\frac{k+1}{k+3}$$

Problem 4

---

a

To calculate the mean, we know that the mean is, for the limits of this given function

$E[X] = \int\limits_0^\infty xf(x) = \frac{1}{3.3} \int\limits_0^\infty xe^{\frac{-x}{3.3}}$

We can integrate by parts using - that is, we know the following to be true:

$\int u\,dv = uv - \int v\,du$

so if we set $u = x$, $du = dx$, $v = -3.3e^{\frac{-x}{3.3}}$, and $dv = e^{\frac{-x}{3.3}}dx$,

$\int xe^{\frac{-x}{3.3}} = 3.3xe^{\frac{-x}{3.3}} + 3.3 \int e^{\frac{-x}{3.3}}dx = 3.3xe^{\frac{-x}{3.3}} - 3.3^2 e^{\frac{-x}{3.3}}$

Substituting this integral back into the original, we obtain

$xe^{\frac{-x}{3.3}} - 3.3e^{\frac{-x}{3.3}}$

evaluating at its limits of 0 and ∞, and with the knowledge that e^{-x} approaches 0 much faster than x approaches infinity,

$xe^{\frac{-x}{3.3}} - 3.3e^{\frac{-x}{3.3}}\Big|_0^\infty = [0+0] - [0-3.3] = 3.3$ ■

To calculate the variance, we solve the following:

$Var(X) = \int\limits_0^\infty x^2 f(x) - E[X]^2$

But since we know the value of $E[X]$ (and squaring it is trivial) we focus our efforts on the left hand side of the equation:

$\int\limits_0^\infty x^2 f(x) = \frac{1}{3.3} \int\limits_0^\infty x^2 e^{\frac{-x}{3.3}}$

As before, we begin to integrate by parts, with $u = x^2$, $du = 2xdx$, $v = -3.3e^{\frac{-x}{3.3}}$, and $dv = e^{\frac{-x}{3.3}}$:

$\frac{1}{3.3}[-3.3x^2 e^{\frac{-x}{3.3}} + 6.6 \int e^{\frac{-x}{3.3}}xdx]$

We integrate by parts again, this time where $u = x$, $du = dx$, $v = -3.3e^{\frac{-x}{3.3}}$, and $dv = e^{\frac{-x}{3.3}}dx$,

$\frac{1}{3.3}\Big[-3.3x^2 e^{\frac{-x}{3.3}} + 6.6[-3.3xe^{\frac{-x}{3.3}} + 3.3 \int e^{\frac{-x}{3.3}}dx]\Big]$

Which is equivalent to

$\frac{1}{3.3}\Big[-3.3x^2 e^{\frac{-x}{3.3}} + 6.6[-3.3xe^{\frac{-x}{3.3}} - 3.3^2 e^{\frac{-x}{3.3}}]\Big]$

As before, evaluating at the limits 0 to ∞ and substitute back in $E[X]^2$

$Var(X) = 2(3.3)^2 - 3.3^2 = 3.3^2 = 10.89$ ■

b

If we substitute 3.3 for β in our previous solutions, we find

$E[X] = xe^{\frac{-x}{\beta}} - \beta e^{\frac{-x}{\beta}}\Big|_0^\infty = [0+0] - [0-\beta] = \beta$
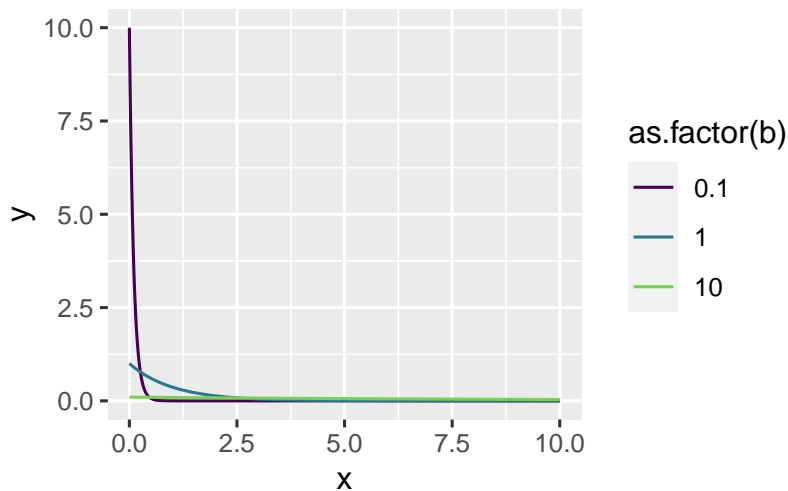
and
$$Var(X) = \frac{1}{\beta}\left[-\beta x^2 e^{\frac{-x}{\beta}} + 2\beta[-\beta x e^{\frac{-x}{\beta}} - \beta^2 e^{\frac{-x}{\beta}}]\right] - \beta^2 = \beta^2$$

c

_____

```
tib <- expand_grid(x = seq(0, 10, by = 0.01), b = c(0.1, 1, 10)) %>% mutate(y = (1/b) *
    exp(-x/b))
```

```
ggplot(tib, aes(x = x, y = y, color = as.factor(b))) + scale_color_viridis_d(end = 0.8) +
    geom_line()
```


Problem 5

_____

a

_____

We know the mean to be
$$\int_0^\infty x f(x)dx = \int_0^\infty \frac{1}{\beta^\alpha \Gamma(\alpha)} x^\alpha e^{\frac{-x}{\beta}}\,dx$$
We begin by substituting, where $u = \frac{x}{\beta}$ and $du = \frac{dx}{\beta}$, which implies
$x = \beta u$ and $dx = \beta du$:
$$\int_0^\infty \frac{1}{\beta^\alpha \Gamma(\alpha)}(\beta u)^\alpha e^{-u}\beta du = \int_0^\infty \frac{1}{\Gamma(\alpha)}u^\alpha e^{-u}\beta du = \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)}\beta$$
Since we know that $\Gamma(\alpha) = (\alpha-1)!$, then
$$\frac{\Gamma(\alpha+1)}{\Gamma(\alpha)}\beta = \alpha\beta \qquad\qquad\qquad \blacksquare$$
For the variance, we know
$$Var(X) = E[X^2] - E[X]^2$$

We begin to solve for the first portion:

$E[X^2] = \int\limits_0^\infty x^2 f(x) = \int\limits_0^\infty \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha+1} e^{\frac{-x}{\beta}} dx$

Using the same substitution as above, we can write it as

$\int\limits_0^\infty \frac{1}{\beta^\alpha \Gamma(\alpha)} (\beta u)^{\alpha+1} e^u \beta du \ = \ \int\limits_0^\infty \frac{1}{\Gamma(\alpha)} \beta^2 u^{\alpha+1} e^u du \ = \ \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)} \beta^2 \ = \ (\alpha+1)\alpha\beta^2 = \alpha^2\beta^2 + \alpha\beta^2$

Solving for second portion is as simple as squaring the value we found for the mean, so in total,

$Var(X) = E[X^2] - E[X]^2 = \alpha^2\beta^2 + \alpha\beta^2 - \alpha^2\beta^2 = \alpha\beta^2$ ∎

b

_____

We can simply substitute in these values for α and β:

$\mu = \alpha\beta = \frac{2p}{2} = p$

$\sigma^2 = \alpha\beta^2 = \frac{4p}{2} = 2p$

Problem 6

_____

a

_____

$E[X] = \int\limits_0^1 x f(x) = \int\limits_0^1 \frac{x^\alpha (1-x)^{\beta-1}}{B(\alpha,\beta)}$

since we know

$B(\alpha,\beta) = \int\limits_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

then

$B(\alpha+1,\beta) = \int\limits_0^1 x^\alpha (1-x)^{\beta-1} dx = \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+1+\beta)}$

so

$\int\limits_0^1 \frac{x^\alpha (1-x)^{\beta-1}}{B(\alpha,\beta)} = \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+1+\beta)} \times \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} = \frac{\alpha\Gamma(\alpha)\Gamma(\beta)}{(\alpha+\beta)\Gamma(\alpha+\beta)} \times \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} = \frac{\alpha}{\alpha+\beta}$ ∎

b

_____

For the variance, like above we solve in portions, starting with the first portion:

$E[X^2] = \int\limits_0^1 x^2 f(x) = \int\limits_0^1 \frac{x^{\alpha+1}(1-x)^{\beta-1}}{B(\alpha,\beta)} = \frac{\Gamma(\alpha+2)\Gamma(\beta)}{\Gamma(\alpha+2+\beta)} \times \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} = \frac{(\alpha+1)(\alpha)\Gamma(\alpha)\Gamma(\beta)}{(\alpha+\beta+1)(\alpha+\beta)\Gamma(\alpha+\beta)} \times \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} = \frac{\alpha(\alpha+1)}{(\alpha+\beta+1)(\alpha+\beta)}$

The second portion is simply the mean squared, so the variance becomes

$$Var(X) = \frac{\alpha(\alpha+1)}{(\alpha+\beta+1)(\alpha+\beta)} - \left(\frac{\alpha}{\alpha+\beta}\right)^2 = \frac{\alpha(\alpha+1)(\alpha+\beta)-(\alpha+\beta+1)\alpha^2}{(\alpha+\beta+1)(\alpha+\beta)^2} =$$
$$\frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$$
∎

Problem 7

---

a

The mean for discrete functions can be written as
$$E[X] = \sum_{x=0}^{\infty} xp(x) = \sum_{x=0}^{\infty} \frac{xe^{-\lambda}\lambda^x}{x!}$$
We know $e^{-\lambda}$ is constant, so we can pull it out of the sum:
$$e^{-\lambda}\sum_{x=0}^{\infty}\frac{x\lambda^x}{x!}$$
We further note that the first term is 0 - so we can safely start the sum at $x=1$:
$$e^{-\lambda}\sum_{x=1}^{\infty}\frac{x\lambda^x}{x!}$$
We now note two things: Each term in our summation has at least one λ, which can be factored out. Additionally, $\frac{x}{x!} = \frac{1}{(x-1)!}$, so the sum can be rewritten as
$$e^{-\lambda}\lambda\sum_{x=1}^{\infty}\frac{\lambda^{x-1}}{(x-1)!}$$
By performing a change of variable, where $y = x-1$, we can more easily see a familiar infinite sum:
$$e^{-\lambda}\lambda\sum_{y=1}^{\infty}\frac{\lambda^y}{y!} = e^{-\lambda}e^{\lambda}\lambda = \lambda$$
∎

b

For this problem, we are given a hint of $E[X(X-1)]$. Let's evaluate this:
$$E[X(X-1)] = \sum_{x=0}^{\infty}\frac{x(x-1)e^{-\lambda}\lambda^x}{x!}$$
We initially note that the first two terms will be 0, so we can shift our sum to start at 2:
$$\sum_{x=2}^{\infty}\frac{x(x-1)e^{-\lambda}\lambda^x}{x!}$$
We also note again that $e^{-\lambda}$ is constant and can be factored out:
$$e^{-\lambda}\sum_{x=2}^{\infty}\frac{x(x-1)\lambda^x}{x!}$$
Further still, we note that the first two terms of $x!$ can be cancelled out with the terms in the numerator:
$$e^{-\lambda}\sum_{x=2}^{\infty}\frac{\lambda^x}{(x-2)!}$$

If we examine the first few terms, we will notice that each term has at least λ^2. We can factor this out as well:

$e^{-\lambda}\lambda^2 \sum\limits_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!}$

A change of variable (where $y = x - 2$) lets us see that this is a familiar sum:

$e^{-\lambda}\lambda^2 \sum\limits_{y=0}^{\infty} \frac{\lambda^y}{(y)!} = e^{-\lambda}e^{\lambda}\lambda^2 = \lambda^2$

We know this to be $E[X(X-1)]$. Manipulating this expression, we see

$E[X(X-1)] = E[X^2 - X]$

Since we know expectation values can be separated, this is equivalent to

$E[X^2] - E[X]$

and we know $E[X] = \lambda$ So we can solve for $E[X^2]$:

$\lambda^2 = E[X^2] - \lambda \implies E[X^2] = \lambda^2 + \lambda$

So,

$Var(X) = \lambda^2 + \lambda - E[X]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$ ∎

Problem 8

---

a

We can write the uniform distribution as

$f(x) = 1$

So the expected value (for $0 < x < 1$) can be calculated as

$E[X] = \int\limits_0^1 x dx = \frac{x^2}{2}\Big|_0^1 = \frac{1}{2}$

Converting to a percentage, the expected percentage of a randomly selected person's body that is covered in freckles is 50%

b

The variance is given by

$Var(X) = E[X^2] - E[X]^2$

Beginning with the first portion,

$E[X^2] = \int\limits_0^1 x^2 dx = \frac{x^3}{3}\Big|_0^1 = \frac{1}{3}$

So,

$Var(X) = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$

Problem 9

_____

a

_____

If we consider X to be the number of songs listened to AFTER the repeated song was listened to, then we know that $P(X = 1) = 0$ - listening to just one song cannot be a repeat of anything.

If we manage to listen to all the songs without a repeat, the next song we listen to MUST be a repeat: there is no song in the list that exists that we have not yet heard $P(X = 1001) = 1$. Any additional songs are not considered. So, $2 \leq X \leq 1001$.

Let's think about what the probability of listening to 3 songs (and hearing a repeat on the third song) is.

The probability of the first song not being a repeat is quite good, so the probability of this event happening is 1. The probability of the seoncd song not being a repeat is also quite good, but one of the songs is forbidden. The probability of having picked a song not yet played before is therefore 999/1000. The third song is where we picked a song we have already heard before. The probability of that happening is quite slim - 2/1000. Altogether, $P(X = 3) = \frac{1000}{1000} \frac{999}{1000} \frac{2}{1000}$

This is all well and good, but suppose we want to generalize this equation for x. Say we listened to x songs, where the xth song was the duplicate. To generalize this, we can write

$P(X = x) = \frac{1000}{1000} \frac{999}{1000} ... \frac{1000-(x+2)}{1000} \frac{x-1}{1000}$

The denominator can be simplified to 1000^x, and the top reeks of a factorial... to me it looks very much like the front end of a factorial. And the front end of a factorial is simply a factorial with the back end divided out. So the numerator is something like

$1000 * 999 * 998 * ... * 1000 - (x + 2) = 1000!/(1000 - (x + 1))!$

So altogether it becomes

$P(X = x) = \frac{1000!(x-1)}{1000^x(1000-(x+1))!}$

b

_____