STAT-36700 Homework 4 - Solutions

Fall 2018

September 28, 2018

This contains solutions for Homework 4. Please note that we have included several additional comments and approaches to the problems to give you better insight.

Problem 1. *Suppose that $X_1, \ldots, X_n \sim \text{Geom}(p)$, i.e. the samples have a geometric distribution with parameter p. A geometric distribution is the distribution of the number of coin flips needed to see one head.*

(a) *Write down the likelihood as a function of the observed data $X_1, \ldots, X_n$, and the unknown parameter p.*

(b) *Compute the MLE of p. In order to do this you need to find a zero of the derivative of the likelihood, and also check that the second derivative of the likelihood at the point is negative.*

(c) *Compute the method-of-moments estimator for p. Is this the same as the MLE?*

(d) ***Extra Credit:*** *Are the estimators you have derived above unbiased? As a hint: think about using Jensen's inequality, and when Jensen's inequality is a strict inequality.*

Solution 1. *We derive each of the results below:*

(a) *We claim the $L(p \mid x_1, x_2, \ldots, x_n) = p^n (1-p)^{\sum_{i=1}^n x_i - n}$*

Proof. We firstly note that $X \sim \mathsf{Geom}(p)$ its PDF is given by $\mathbb{P}(X = k) = (1-p)^{k-1} p$. Now we have the likelihood as:

$$\begin{aligned} L(p \mid x_1, x_2, \ldots, x_n) &= \prod_{i=1}^n \mathbb{P}(X_i = x_i) \\ &= \prod_{i=1}^n (1-p)^{x_i - 1} p \\ &= p^n (1-p)^{\sum_{i=1}^n x_i - n} \end{aligned}$$

□

(b) *We claim that $\hat{p}_{MLE} = \frac{1}{\overline{X}_n}$*

Proof. From the previous part we have that the Likelihood of p is given by:

$$L(p \mid x_1, x_2, \ldots, x_n) = p^n (1-p)^{\sum_{i=1}^n x_i - n}$$

. We can then derive the log-likelihood of p as follows:

$$L(p \mid x_1, x_2, \ldots, x_n) = p^n (1-p)^{\sum_{i=1}^n x_i - n}$$
$$\implies l(p \mid x_1, x_2, \ldots, x_n) = n \log(p) + \left(\sum_{i=1}^n x_i - n\right) \log(1-p) \quad \text{(we are now maximizing the log-likelihood)}$$

Now we can differentiate the log-likelihood and set to 0 to find the MLE as follows:

$$\begin{aligned}\frac{\partial l}{\partial p} = \frac{n}{p} - \frac{\sum_{i=1}^n x_i - n}{1-p} = 0 \quad &\text{(setting partial derivative to 0)} \\ \implies n - np = p\sum_{i=1}^n x_i - np& \\ \implies n = p\sum_{i=1}^n x_i& \\ \implies p = \frac{n}{\sum_{i=1}^n x_i}& \\ = \frac{1}{\overline{x}_n}&\end{aligned}$$

We then verify that this is a maximum using the second derivative as follows:

$$\begin{aligned}\frac{\partial^2 l}{\partial p^2} &= -\frac{n}{p^2} - \frac{n - \sum_{i=1}^n x_i}{(1-p)^2} \\ &= -\left[\frac{n}{p^2} + \frac{\sum_{i=1}^n x_i - n}{(1-p)^2}\right] \\ &< 0\end{aligned}$$

Since it has negative curvature for all p we have that $\hat{p}_{MLE} = \frac{1}{\overline{X}_n}$ as required. □

(c) *We claim that* $\hat{p}_{MOM} = \frac{1}{\overline{X}_n}$

Proof. [1]For $X \sim \text{Geom}(p)$ we have that $\mathbb{E}(X) = \frac{1}{p}$.
Now we use the method of moments as follows:

[1] **Note:** Prove this as an exercise

$$\begin{aligned}\frac{1}{\hat{p}_{MOM}} &= \frac{1}{n}\sum_{i=1}^n x_i \\ &= \overline{x}_n \\ \implies \hat{p}_{MOM} &= \frac{1}{\overline{x}_n}\end{aligned}$$

So the method of moments is $\hat{p}_{MOM} = \frac{1}{\overline{X}_n}$ which is equal to $\hat{p}_{MLE}$ in this case. □

(d) ***Extra Credit:*** *We claim that* $\hat{p}_{MLE} = \hat{p}_{MOM} = \frac{1}{\overline{X}_n}$ *are biased estimators.*

Proof. Firstly we note that

$$\begin{aligned}
\mathbb{E}(\overline{X}_n) &= \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \mathbb{E}(X_1) \quad \text{(using linearity of expectation and IID of } X_i\text{'s)} \\
&= \frac{1}{p} \quad \text{(Using expectation of } \mathsf{Geom}p \text{ variable (Exercise: derive this))}
\end{aligned}$$

Now per the hint, using Jensen's inequality for the convex function[2] $x \mapsto \frac{1}{x}$ we have:

[2] This is checked by noting the second derivative $\frac{\partial^2}{\partial x^2}\frac{1}{x} = \frac{1}{x^3} > 0 \quad \forall x > 0$

$$\begin{aligned}
\mathbb{E}(p_{\hat{MLE}}) &= \mathbb{E}\left(\frac{1}{\overline{X}_n}\right) \\
&> \frac{1}{\mathbb{E}(\overline{X}_n)} \quad \text{(note the strict inequality since our convex map is not linear)} \\
&= \frac{1}{\frac{1}{p}} \\
&= p
\end{aligned}$$

Since $\mathbb{E}(p_{\hat{MLE}}) = \mathbb{E}(p_{\hat{MOM}}) > p$ our estimators are biased. □

Problem 2. *Suppose we have samples $X_1, \ldots, X_n \sim \mathsf{Unif}[0, \theta]$.*

(a) *Write down the likelihood as a function of the observed data $X_1, \ldots, X_n$, and the unknown parameter θ.*

(b) *Compute the MLE of θ.*

(c) *Use the method of moments to derive an estimator of θ. Is this the same as the MLE?*

Solution 2. *We derive each of the results below:*

(a) *We claim the $L(\theta \mid x_1, x_2, \ldots, x_n) =$*

Proof.

$$\begin{aligned}
L(\theta \mid x_1, x_2, \ldots, x_n) &= \prod_{i=1}^{n} \frac{1}{\theta}\mathbb{1}(0 \le x_i \le \theta) \\
&= \left(\frac{1}{\theta}\right)^n \prod_{i=1}^{n} \mathbb{1}(0 \le x_i \le \theta) \\
&= \left(\frac{1}{\theta}\right)^n \mathbb{1}(0 \le x_{(1)} \le x_{(n)} \le \theta)
\end{aligned}$$

□

(b) [3] *We claim that $\hat{\theta}_{MLE} = X_{(n)}$*

[3] **Note:** In this case it is critical to understand that the parameter we are seeking to estimate i.e. θ is in the **support** of our likelihood. As such we can't just differentiate the log-likelihood per usual and have to carefully consider the shape of the likelihood over it's support and choose the θ value that maximizes it.

Proof. We note that from part (a) that the log likelihood is:

$$l(\theta \mid x_1, x_2, \ldots, x_n) = -n \log(\theta) \mathbb{1}(0 \leq x_{(1)} \leq x_{(n)} \leq \theta)$$
$$\implies \frac{\partial l}{\partial \theta} = -\frac{n}{\theta} \mathbb{1}(0 \leq x_{(1)} \leq x_{(n)} \leq \theta)$$

which is a decreasing function of θ over it's support. This function is then maximized over it's support when $\theta = x_{(n)}$. We then have that $\hat{\theta}_{MLE} = X_{(n)}$ as required. □

(c) *We claim that* $\hat{\theta}_{MOM} = 2\overline{X}_n$

Proof.

$$\begin{aligned}
\mathbb{E}(X) &= \int_0^\theta \frac{x}{\theta} dx \\
&= \left[\frac{x^2}{2\theta}\right]_0^\theta \\
&= \frac{\theta^2}{2\theta} \\
&= \frac{\theta}{2} \\
\implies \frac{\hat{\theta}_{MOM}}{2} &= \frac{1}{n}\sum_{i=1}^n x_i \\
\implies \hat{\theta}_{MOM} &= 2\overline{x}_n
\end{aligned}$$

We then have $\hat{\theta}_{MOM} = 2\overline{X}_n$ as required. This is not necessarily the same as the MLE. □

Problem 3. *In this question we will explore what is known as Monte Carlo integration. We often wish to calculate an integral of the form:*

$$I(f) = \int_0^1 f(x) dx,$$

but cannot do so analytically. The Monte Carlo method is a way to approximate this integral.

(a) *Re-express this integral as an expected value of the function* f*. What is the distribution of the underlying random variable.*

(b) *The above suggests that we can generate random variables in a certain way and approximate the integral of interest. Use the weak law of large numbers (WLLN) to give a precise (asymptotic) guarantee for this method.*

(c) *As a concrete example, we consider the evaluation of:*

$$I(f) = \frac{1}{\sqrt{2\pi}} \int_0^1 \exp(-x^2/2)dx.$$

Use R (or any other programming language) to approximate the above integral using 1000 random samples. Report your result. Repeat this 100 times just to get a sense of the variability - you do not need to report these 100 numbers.

(d) *Use the standard Gaussian CDF to give a precise evaluation of this integral. You should use R (or any other language) to evaluate the CDF.*

Solution 3. *We derive each of the results below:*

(a) *We claim that this is $\mathbb{E}_X(f(X))$ where $X \sim \mathsf{Unif}[0,1]$.*

Proof. For $X \sim \mathsf{Unif}[0,1]$ we have the PDF as:

$$p_X(x) = \mathbb{1}(0 \le x \le 1)$$

. So we have:

$$\begin{aligned}
\mathbb{E}_X(f(X)) &:= \int_{\mathcal{X}} f(x) p_X(x) dx \\
&= \int_{\mathcal{X}} f(x) \mathbb{1}(0 \le x \le 1) dx \\
&= \int_0^1 f(x) dx
\end{aligned}$$

as required □

(b) *Now that we have expressed the required integral as the expectation of a random variable $X \sim \mathsf{Unif}[0,1]$, we have that for $X_1, X_2, \ldots X_n \sim X \sim \mathsf{Unif}[0,1]$ by the WLLN that:*

$$\frac{1}{n} \sum_{i=1}^{n} f(X_i) \xrightarrow{p} \mathbb{E}_X(f(X)) = \int_0^1 f(x) dx$$

(c) *As a concrete example, we consider the evaluation of:*

$$I(f) = \frac{1}{\sqrt{2\pi}} \int_0^1 \exp(-x^2/2)dx.$$

We can use R *to approximate the above integral using 1000 random samples. Report your result.*

```
# install.packages("tidyverse")
library(tidyverse)
```

```r
# Set seed for reproducibility
set.seed(35842)

# Define the required function we want to integrate
f <- function(x){
    base::return((1/sqrt(2*pi))*exp(-(x^2/2)))
}

# Define the Monte Carlo simulation
monte_carlo <- function(inp_fn, n){
    # Generate random samples
    samps <- purrr::map_dbl(.x = 1:n, ~ runif(n = 1,
                                              min = 0,
                                              max = 1))

    # Approximate the integral
    integl <- mean(inp_fn(samps))

    base::return(integl)
}

# Single run
one_sim <- monte_carlo(inp_fn = f, n = 1000)
one_sim
```

This returns a value of 0.3416494.

We can repeat this 100 times as follows:

```r
# Let's replicate the simulation 100 times
num_replications <- 100
sims <- replicate(expr = monte_carlo(inp_fn = f, n = 1000),
                  n = num_replications)

# Let's plot a histogram
hist(sims)
```

(d) *We can represent this exactly as the difference of CDF of the standard normal distribution as follows:*

$$
\begin{aligned}
I(f) &= \frac{1}{\sqrt{2\pi}} \int_0^1 \exp(-x^2/2)dx. \\
&= \int_{-\infty}^1 \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)dx. - \int_{-\infty}^0 \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)dx. \\
&= \Phi(1) - \Phi(0)
\end{aligned}
$$

This can be numerically calculated (approximated) in R *using the following command:*

```
pnorm(1) - pnorm(0)
```

which evaluates to 0.3413447.

Problem 4. *The Poisson distribution with parameter λ, has pmf:*

$$\mathbb{P}(X = k) = \frac{\exp(-\lambda)\lambda^k}{k!}.$$

For large values of λ the Poisson distribution is well approximated by a Gaussian distribution.

(a) *Use moment matching to find the Gaussian that best approximates a $\text{Poi}(\lambda)$ distribution. In other words, we can use $N(\mu, \sigma^2)$ to approximate the Poisson, of we choose μ and σ^2 to match the mean and variance of the Poisson.*

(b) *Suppose that we have a system that emits a random variable X particles according to a Poisson distribution with mean $\lambda = 900$ per hour. Use the above approximation to calculate the probability $\mathbb{P}(X > 950)$. You should express this in terms of an appropriate standard Gaussian quantile, i.e., express your answer in terms of the function $\Phi(z) = \mathbb{P}(Z \leq z)$ where Z has a standard normal distribution.*

(c) *Use R to compute the value of $\mathbb{P}(X > 950)$, approximately using the Gaussian quantile and exactly using the Poisson CDF. There are functions built in to R for each of these.*

Solution 4. *We derive each of the results below:*

(a) *We are given that $X \sim \mathsf{Poiss}(\lambda)$ and $Y \sim \mathcal{N}(\mu, \sigma^2)$. We know that $\mathbb{E}(X^2) = \mathsf{Var}(X) + (\mathbb{E}(X))^2$.*
Equating first moments we have:

$$\begin{aligned} \mathbb{E}(X) &= \mathbb{E}(Y) \\ \implies \mu &= \lambda \end{aligned}$$

Equating second moments we have

$$\begin{aligned} \mathbb{E}\left(X^2\right) &= \mathbb{E}\left(Y^2\right) \\ \implies \lambda^2 + \lambda &= \sigma^2 + \mu^2 \\ &= \sigma^2 + \lambda^2 \\ \implies \sigma^2 &= \lambda \end{aligned}$$

So the closest Gaussian approximation to the Poisson by matching the first 2 moments is $Y \sim \mathcal{N}(\lambda, \lambda)$.

(b) *We have from part (a) that* $X \sim Y \sim \mathcal{N}(\mu = \lambda, \sigma^2 = \lambda)$ *(approximately). This shows that:*

$$\begin{aligned}
\mathbb{P}(X > 950) &= 1 - \mathbb{P}(X \leq 950) \\
&= 1 - \mathbb{P}\left(\frac{X-\mu}{\sigma} \leq \frac{950-\mu}{\sigma}\right) \quad = 1 - \mathbb{P}\left(\frac{X-\lambda}{\sqrt{\lambda}} \leq \frac{950-\lambda}{\sqrt{\lambda}}\right) \\
&= 1 - \mathbb{P}\left(Z \leq \frac{50}{30}\right) \\
&= 1 - \Phi\left(\frac{5}{3}\right)
\end{aligned}$$

(c) *We can use* R *to perform the approximation as follows:*

```
# Setup parameters
lambd <- 900
mu <- lambd
sigma_sq <- lambd
t_val <- 950
t_val_std <- (t_val - mu)/sqrt(sigma_sq)

# Normal approximation to the poisson
1 - pnorm(t_val_std)
```

which gives a value of 0.04779035.

If we calculate using the in-built Poisson functions we get the more precise numerical estimation as:

```
1 - ppois(q = 950, lambda = lambd)
```

which gives a value of 0.04711902.

Problem 5. *A Binomial RV is the sum of independent Bernoullis and is thus well approximated by a Gaussian by the central limit theorem. We will use this to perform a rudimentary hypothesis test.*

We believe a coin is fair and toss it 100 times. It lands heads up 60 times. Use moment matching (or the CLT) to approximate the probability $\mathbb{P}(X \geq 60)$ *assuming the coin is fair in terms of the standard Gaussian CDF.*

Use R to give a numerical value. Roughly, our intuition is that if this value is small then we should be "suspicious" of our initial hypothesis that the coin is fair. Are you suspicious?

Solution 5. *Let* $Y_1, ..., Y_n \sim$ *Bernoulli*(p) *with* $p = 1/2$. *Then* $X = \{\text{number of times landing heads up}\} = \sum_{i=1}^{n} Y_i$. *Since* $n = 100$ *is large enough that by applying the CLT, we have* $\frac{\sqrt{n}(\bar{Y}-\mu)}{\sigma} \sim \mathcal{N}(0,1)$.

Equivalently,

$$\frac{\sqrt{n}(\bar{Y}-\mu)}{\sigma} \sim \mathcal{N}(0,1) \Rightarrow \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$$
$$\Rightarrow \sum_{i=1}^{n} Y_i \sim \mathcal{N}(n\mu, n\sigma^2)$$

Since $\mu = \mathbb{E}[Y_i] = p = 1/2$, $\sigma^2 = Var(Y_i) = p(1-p)$, *we have* $X = \sum_{i=1}^{n} Y_i \sim \mathcal{N}(np, np(1-p)) = \mathcal{N}(50, 25)$. *Hence*

$$\begin{aligned}\mathbb{P}(X \geq 60) &= \mathbb{P}(\frac{X-50}{5} \geq \frac{60-50}{5})\\ &= \mathbb{P}(Z \geq 2)\\ &= 1 - F_Z(2)\\ &\approx 0.02275 \quad (\textit{can be computed using } \texttt{1-pnorm(2)} \textit{ in } \texttt{R})\end{aligned}$$

Where Z is a standard Gaussian random variable. This value is small enough that we should be suspicious.

Problem 6. *Suppose that* $X_1, \ldots, X_n$ *are repeated measurements of a quantity* μ, *and that* $\mathbb{E}[X_i] = \mu$, *and* $\mathrm{Var}(X_i) = \sigma^2$. *Further, let us suppose that each* $X_i \in [0,1]$. *Let* $\overline{X}_n$ *denote the average of the measurements.*

(a) *Use the fact that each* $X_i \in [0,1]$ *to give some bounds on* μ *and* σ.

(b) *Suppose that we take 16 measurements and that* $\sigma^2 = \frac{1}{12}$. *Use the CLT to approximate the probability that the average deviates from* μ *by more than* 0.5.

(c) *Use Chebyshev's inequality to give an upper bound on the same quantity.*

(d) *Repeat the above calculation but now use Hoeffding's inequality.*

(e) *Now use R to estimate this probability in the following way. Suppose that each* X_i *is* $U[0,1]$. *The mean is* 0.5 *and the variance is exactly* $1/12$. *Draw 16 measurements and track if the sample mean is within 0.5 of the true mean. Repeat this 1000 times to get an accurate estimate. Compare the answer to what you obtained analytically. Particularly, order the confidence intervals by length.*

Solution 6.(a) $0 \leq X_i \leq 1 \Rightarrow 0 \leq \mathbb{E}[X_i] \leq 1 \Rightarrow 0 \leq \mu \leq 1$.
$0 \leq Var(X_i) = E[(X_i - \mathbb{E}[X_i])^2] \leq (1-0)^2 \leq 1 \Rightarrow 0 \leq \sigma \leq 1$ *(Note that we can actually have* $Var(X_i) \leq 1/4$ *by the Popoviciu's inequality).*

(b) *By the CLT,* $\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \sim \mathcal{N}(0,1)$*, then*

$$\begin{aligned}
\mathbb{P}(|\bar{X}-\mu| > 0.5) &= \mathbb{P}(\frac{\sqrt{n}|\bar{X}-\mu|}{\sigma} > \frac{0.5\sqrt{n}}{\sigma}) \\
&= \mathbb{P}(|Z| > \frac{0.5 \times \sqrt{16}}{\sqrt{1/12}}) \\
&= \mathbb{P}(|Z| > 4\sqrt{3}) \\
&= 2 \times \mathbb{P}(Z > 4\sqrt{3}) \\
&= 4.26 \times 10^{-12} \quad (\text{R command: } \texttt{2*(1-pnorm(4*sqrt(3)))})
\end{aligned}$$

(c) *First notice that* $\mathbb{E}(\bar{X}) = \mu$ *and* $Var(\bar{X}) = Var(X_i)/n = \sigma^2/n$*. Then by Chebyshev's inequality,*

$$\begin{aligned}
\mathbb{P}(|\bar{X}-\mu| > 0.5) &\leq \frac{Var(\bar{X})}{0.5^2} \\
&= \frac{\sigma^2}{0.25n} \\
&= \frac{1}{12 \times 0.25 \times 16} \\
&= \frac{1}{48} \approx 0.0208
\end{aligned}$$

(d) *By Hoeffding's inequality,*

$$\begin{aligned}
\mathbb{P}(|\bar{X}-\mu| > 0.5) &\leq 2\exp(-\frac{2n^2 0.5^2}{\sum_{i=1}^{n}(1-0)^2}) \\
&= 2\exp(-2 \times 0.25 \times 16) \\
&= 0.00067
\end{aligned}$$

(e) *Use* R *to simulate the samples:*

```
# initialize
L = rep(0, 1000)

for(k in 1:1000){
  # generate 16 uniformly distributed samples
  X = runif(16, min = 0, max = 1)
  # sample mean
  X.mean = mean(X)
  # check if the mean is within 0.5 of the true mean
  L[k] = (abs(X.mean - 0.5) > 0.5)
}
# probability = 0
prob = sum(L)/length(L)
```

Based on the R *simulations,* $\mathbb{P}(|\bar{X} - \mu| > 0.5) = 0$*. Summarizing parts (b)(c)(d)(e), we have*

$$\mathbb{P}_R < \mathbb{P}_{CLT} < \mathbb{P}_{Hoeffding's} < \mathbb{P}_{Chebyshev's}$$

and hence, the α*-level confidence intervals for (b)(c)(d)(e) satisfy*

$$CI_R \subset CI_{CLT} \subset CI_{Hoeffding's} \subset CI_{Chebyshev's}$$

Problem 7. Conjugate Priors: *In lectures we have seen that the Beta and Binomial are conjugate distributions, i.e. if we are estimating a Binomial parameter and use a Beta prior then the posterior is a Beta distribution. There is a similar relationship between Gamma and Poisson distributions. You can use any reference (Wikipedia) for the distributions you need - for the Gamma use the shape/rate parameterization.*

(a) *Suppose the data* $X_1, \ldots, X_n \sim \text{Poi}(\lambda)$*. Write down the likelihood as a function of the parameter* λ*.*

(b) *Compute the MLE for* λ*.*

(c) *Assume that* $\lambda \sim \text{Gamma}(\alpha, \beta)$*, and write down the posterior distribution over the parameter* λ*. Show that this posterior distribution is a Gamma distribution, and compute its parameters.*

(d) *The mean of a Gamma distribution is* α/β*. Compute the posterior mean. This will be our point estimate.*

(e) *Write the posterior mean as a convex combination of the prior mean and the MLE. What happens if* α, β *are fixed and* $n \to \infty$*?*

Solution 7.*(a)*

$$\begin{aligned}
L(\lambda; x_1, ..., x_n) = L(\lambda) &= \prod_{i=1}^{n} \mathbb{P}(X_i = x_i) \\
&= \prod_{i=1}^{n} \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} \\
&= \frac{e^{-n\lambda} \lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}
\end{aligned}$$

(b) *Since* $L(\lambda) \geq 0$*,* $\lambda_{MLE} = \arg\max_{\lambda} L(\lambda) = \arg\max_{\lambda} \log L(\lambda)$*.*

Such a λ *satisfies* $\frac{\partial \log L(\lambda)}{\partial \lambda} = 0$ *and* $\frac{\partial^2 \log L(\lambda)}{\partial \lambda^2} \leq 0$*.*

$$\log L(\lambda) = -n\lambda + \sum_{i=1}^{n} x_i \log \lambda - \log \prod_{i=1}^{n} x_i!$$

$$\frac{\partial \log L(\lambda)}{\partial \lambda} = -n + \frac{\sum_{i=1}^{n} x_i}{\lambda} = 0$$
$$\Rightarrow \lambda = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\frac{\partial^2 \log L(\lambda)}{\partial \lambda^2} = -\frac{\sum_{i=1}^{n} x_i}{\lambda^2} < 0.$$

Therefore $\lambda_{MLE} = \frac{\sum_{i=1}^{n} x_i}{n}$.

(c) *From lecture notes 9, we have*

$$\begin{aligned} p(\lambda|x_1, ..., x_n) &\propto L(\lambda)p(\lambda) \\ &\propto \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} \\ &\propto \frac{(\beta+n)^{\alpha+\sum_{i=1}^{n} x_i}}{\Gamma(\alpha+\sum_{i=1}^{n} x_i)} \lambda^{\alpha+\sum_{i=1}^{n} x_i - 1} e^{-(\beta+n)\lambda} \end{aligned}$$

Therefore $\lambda|X_1, ..., X_n \sim Gamma(\alpha + \sum_{i=1}^{n} X_i, \beta + n)$.

(d) *This is*

$$\lambda_{Bayes} = \frac{\alpha + \sum_{i=1}^{n} x_i}{\beta + n}$$

(e)

$$\lambda_{Bayes} = \frac{\beta}{\beta+n}\frac{\alpha}{\beta} + \frac{n}{\beta+n}\frac{\sum_{i=1}^{n} x_i}{n}$$

when $n \to \infty$, $\lambda_{Bayes} = \lambda_{MLE}$.

Problem 8. Computing a Bayes rule: *Suppose that* $X \sim N(\mu, 1)$ *is one sample from a Normal Let* $\mu \sim N(0, 1)$ *be a prior. Assume squared error loss. Recall that the Bayes estimator is the estimator that minimizes the expected posterior loss.*

(a) *Compute the posterior distribution for* μ.

(b) *For the squared loss, compute the Bayes estimator.*

(c) **Extra credit:** *Compute the Bayes estimator using the loss* $|\hat{\mu} - \mu|$.

Solution 8.(a) *From lecture notes 9, we have*

$$\begin{aligned} f(\mu|x) &\propto L(\mu;x)f(\mu) \\ &\propto \frac{1}{\sqrt{2\pi}}\exp -\frac{(x-\mu)^2}{2} \times \frac{1}{\sqrt{2\pi}}\exp -\frac{\mu^2}{2} \\ &\propto \exp -\frac{2\mu^2 - 2x\mu + x^2}{2} \\ &= \exp -\frac{(\mu - \frac{x}{2})^2 + \frac{x^2}{4}}{2\cdot(\frac{1}{2})} \\ &\propto \exp -\frac{(\mu - \frac{x}{2})^2}{2\cdot(\frac{1}{2})} \end{aligned}$$

Therefore $\mu|X \sim \mathcal{N}(\frac{X}{2}, \frac{1}{2})$.

(b) *Let* $R_{L2}(\mu,\hat{\mu}) = \mathbb{E}_{\mu|X}(\hat{\mu} - \mu)^2$.
$\hat{\mu}_{L2} = \arg\min_{\mu} R_{L2}(\mu,\hat{\mu})$. *Such a* $\hat{\mu}_{L2}$ *satisfies:* $\frac{\partial R_{L2}(\mu,\hat{\mu})}{\partial \hat{\mu}} = 0$ *and* $\frac{\partial^2 R_{L2}(\mu,\hat{\mu})}{\partial \hat{\mu}^2} \geq 0$.

$$R_{L2}(\mu,\hat{\mu}) = \mathbb{E}_{\mu|X}(\hat{\mu}-\mu)^2 = \int_{-\infty}^{+\infty} (\hat{\mu}-\mu)^2 f_{\mu|X}(\mu) f\mu$$

$$\begin{aligned} \frac{\partial R_{L2}(\mu,\hat{\mu})}{\partial \hat{\mu}} &= \int_{-\infty}^{+\infty} \frac{\partial(\hat{\mu}-\mu)^2}{\partial\hat{\mu}} f_{\mu|X}(\mu) f\mu \\ &= \int_{-\infty}^{+\infty} 2(\hat{\mu}-\mu) f_{\mu|X}(\mu) f\mu = 0 \\ &\Rightarrow \int_{-\infty}^{+\infty} \hat{\mu} f_{\mu|X}(\mu) f\mu = \int_{-\infty}^{+\infty} \mu f_{\mu|X}(\mu) f\mu \\ &\Rightarrow \hat{\mu} \int_{-\infty}^{+\infty} f_{\mu|X}(\mu) f\mu = \mathbb{E}_{\mu|X}(\mu) \\ &\Rightarrow \hat{\mu} = \mathbb{E}_{\mu|X}(\mu) = \frac{X}{2} \end{aligned}$$

And

$$\frac{\partial^2 R_{L2}(\mu,\hat{\mu})}{\partial\hat{\mu}^2} = \int_{-\infty}^{+\infty} 2 f_{\mu|X}(\mu) f\mu = 2 > 0.$$

Therefore $\hat{\mu}_{L2} = \frac{X}{2}$.

(c) *The idea is same as (b), and one can get*[4]

$$R_{L1}(\mu,\hat{\mu}) = \mathbb{E}_{\mu|X}|\hat{\mu}-\mu| = \int_{-\infty}^{+\infty} |\hat{\mu}-\mu| f_{\mu|X}(\mu) f\mu$$

[4] **Note:** For $f(x) = |x|$, the derivative $f'(x) = \begin{cases} 1 & x > 0 \\ -1 & x < 0 \end{cases}$.

$$
\begin{aligned}
\frac{\partial R_{L1}(\mu,\hat{\mu})}{\partial \hat{\mu}} &= \int_{-\infty}^{+\infty} \frac{\partial|\hat{\mu}-\mu|}{\partial \hat{\mu}} f_{\mu|X}(\mu) f\mu \\
&= \int_{-\infty}^{\hat{\mu}} f_{\mu|X}(\mu) f\mu - \int_{\hat{\mu}}^{+\infty} f_{\mu|X}(\mu) f\mu = 0 \\
&\Rightarrow \mathbb{P}(\mu \leq \hat{\mu}) = \mathbb{P}(\mu > \hat{\mu}) \\
&\Rightarrow \hat{\mu} = \frac{X}{2} \qquad \textit{By symmetry of normal distribution}
\end{aligned}
$$

And[5]

$$
\frac{\partial^2 R_{L1}(\mu,\hat{\mu})}{\partial \hat{\mu}^2} = 0.
$$

Therefore $\hat{\mu}_{L1} = \frac{X}{2}$.

[5] **Note:** In this case for the absolute value loss, the Bayes estimator is the median of the posterior, not mean per the squared loss. For the normal distribution the median and the mean coincide so the estimator is the same as the previous part