EE178 HW 8 Solutions

Nov 30, 2016

1. **(15 pts) Gaussian Distribution**
   If a set of grades on a probability examination in an inferior school (not Stanford!) are approximately Gaussian distributed with a mean of 64 and a standard deviation of 7.1, find:

   (a) (8 pts) the lowest passing grade if the bottom 5% of the students fail the class
   Answer: We are told that $\mathbf{P}(X \leq 1.65) \approx 0.95$ when $\mu = 0$ and $\sigma = 1$. As the normal distribution is characterized entirely by μ and σ, we see that, in general, for a normally distributed random variable X with mean μ and standard deviation σ, we have that $\mathbf{P}(X - \mu \leq 1.65\sigma) \approx 0.95$. Since the normal distribution is symmetric about μ, we have that

   $$\begin{aligned} \mathbf{P}(\mu - X \leq 1.65\sigma) &\approx 0.95 \\ \mathbf{P}(X \geq \mu - 1.65\sigma) &\approx 0.95 \\ \mathbf{P}(X \geq 64 - (1.65)(7.1)) &\approx 0.95 \\ \mathbf{P}(X \geq 52.285) &\approx 0.95 \end{aligned}$$

   Therefore, the lowest passing grade is approximately 53.

   (b) (7 pts) the highest B if the top 10% of the students are given A's.
   Answer: Proceeding in similar fashion to part (a), we have

   $$\begin{aligned} \mathbf{P}(X - \mu \leq 1.3\sigma) &\approx 0.9 \\ \mathbf{P}(X \leq \mu + 1.3\sigma) &\approx 0.9 \\ \mathbf{P}(X \leq 64 + (1.3)(7.1)) &\approx 0.9 \\ \mathbf{P}(X \leq 73.23) &\approx 0.9 \end{aligned}$$

   Therefore, the highest B is approximately 73.

2. **(15 pts) Confidence Intervals Again**

   (a) (5 pts) Answer: Using Chebyshev's inequality,

   $$\mathbf{P}\left(|\hat{p} - p| > 0.03\right) \leq \frac{\mathrm{Var}(\hat{p})}{0.03^2} = \frac{p(1-p)}{n \cdot 0.03^2}$$

   We want the right hand side to be ≤ 0.05, i.e.,

   $$\frac{p(1-p)}{n \cdot 0.03^2} \leq 0.05,$$

   $$n \geq \frac{p(1-p)}{0.05 \cdot 0.03^2}.$$

Since $p(1-p) \leq 0.25$, the above can be guaranteed if

$$n \geq \frac{0.25}{0.05 \cdot 0.03^2},$$
$$n \geq 5556,$$

which is more than 500 since more samples are required for a smaller margin of error.

(b) (10 pts) Answer: By central limit theorem, $\frac{\hat{p}-p}{\sqrt{\mathrm{Var}(\hat{p})}}$ tends to the standard Gaussian distribution when the number of samples tends to infinity. Let

$$X = \frac{\hat{p}-p}{\sqrt{\mathrm{Var}(\hat{p})}} = \frac{\hat{p}-p}{\sqrt{\frac{1}{n}p(1-p)}}$$

and $Z \sim N(0,1)$.

$$\begin{aligned}
\mathbf{P}\left(|\hat{p}-p| > 0.03\right) &= \mathbf{P}\left(|X| > \frac{0.03}{\sqrt{\frac{1}{n}p(1-p)}}\right) \\
&\leq \mathbf{P}\left(|X| > \frac{0.03}{\sqrt{\frac{1}{n}\cdot 0.25}}\right) \\
&= \mathbf{P}\left(|X| > 0.06\sqrt{n}\right) \\
&\approx \mathbf{P}\left(|Z| > 0.06\sqrt{n}\right) \\
&= 2\mathbf{P}\left(Z > 0.06\sqrt{n}\right) \\
&\leq 0.05
\end{aligned}$$

when $0.06\sqrt{n} \geq 1.96$ (since $\mathbf{P}(Z > 1.96) \approx 0.025$ from standard normal table), $n \geq 1068$, which is less conservative than part (a).

3. **(30 pts) Convolution and Probability**

(a) (2 pts) Answer: It follows from the total probability rule that

$$\mathbf{P}(V=b) = \sum_a \mathbf{P}(V=b|U=a)\mathbf{P}(U=a).$$

(b) (8 pts) Answer: We have

$$\begin{aligned}
f_V(b) &= \int_{-\infty}^{\infty} f_{UV}(a,b)da \\
&= \int_{-\infty}^{\infty} f_{V|U}(b|a)f_U(a)da.
\end{aligned}$$

The unit on the left hand side is 1/(unit of V). The unit of $f_{V|U}(b|a)f_U(a)$ is 1/((unit of V)(unit of U)). After integrating over the domain of U, the unit of the right hand side is (unit of U)/((unit of V)(unit of U)) = 1/(unit of V).

(c) (8 pts) Answer: If X, Y independent, then by (b),

$$f_{X+Y}(b) = \int_{-\infty}^{\infty} f_{X+Y|X}(b|a)f_X(a)da.$$

If we know $X = a$, then Y still follows the pdf $f_Y(b)$ (since Y is independent of X), so $Y + X = Y + a$ follows the pdf $f_Y(b-a)$. Hence $f_{X+Y|X}(b|a) = f_Y(b-a)$,

$$f_{X+Y}(b) = \int_{-\infty}^{\infty} f_X(a) f_Y(b-a) da.$$

Hence f_{X+Y} is the convolution of f_X and f_Y.

(d) (8 pts) Answer: To simplify the calculations, assume $\mu_X = \mu_Y = 0$. We have

$$\begin{aligned}
f_{X+Y}(b) &= \int_{-\infty}^{\infty} f_X(a) f_Y(b-a) da \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left(-\frac{a^2}{2\sigma_X^2}\right) \cdot \frac{1}{\sqrt{2\pi\sigma_Y^2}} \exp\left(-\frac{(b-a)^2}{2\sigma_Y^2}\right) da \\
&= \frac{1}{2\pi\sigma_X\sigma_Y} \int_{-\infty}^{\infty} \exp\left(-\frac{a^2}{2\sigma_X^2} - \frac{(b-a)^2}{2\sigma_Y^2}\right) da \\
&= \frac{1}{2\pi\sigma_X\sigma_Y} \int_{-\infty}^{\infty} \exp\left(-\left(\frac{1}{2\sigma_X^2} + \frac{1}{2\sigma_Y^2}\right) a^2 + \left(\frac{b}{\sigma_Y^2}\right) a - \frac{b^2}{2\sigma_Y^2}\right) da \\
&= \frac{1}{2\pi\sigma_X\sigma_Y} \int_{-\infty}^{\infty} \exp\left(-\left(\frac{1}{2\sigma_X^2} + \frac{1}{2\sigma_Y^2}\right) a^2 + \left(\frac{b}{\sigma_Y^2}\right) a - \frac{b^2}{2\sigma_Y^2}\right) da \\
&= \frac{1}{\sqrt{2\pi}\sigma_X\sigma_Y\sqrt{\frac{1}{\sigma_X^2} + \frac{1}{\sigma_Y^2}}} \int_{-\infty}^{\infty} \frac{\sqrt{\frac{1}{\sigma_X^2} + \frac{1}{\sigma_Y^2}}}{\sqrt{2\pi}} \\
&\qquad \exp\left(-\left(\frac{1}{2\sigma_X^2} + \frac{1}{2\sigma_Y^2}\right)\left(a - \frac{\frac{b}{2\sigma_Y^2}}{\frac{1}{2\sigma_X^2} + \frac{1}{2\sigma_Y^2}}\right)^2 + \frac{\left(\frac{b}{2\sigma_Y^2}\right)^2}{\frac{1}{2\sigma_X^2} + \frac{1}{2\sigma_Y^2}} - \frac{b^2}{2\sigma_Y^2}\right) da \\
&\overset{(i)}{=} \frac{1}{\sqrt{2\pi}\sqrt{\sigma_X^2 + \sigma_Y^2}} \exp\left(\frac{\left(\frac{b}{2\sigma_Y^2}\right)^2}{\frac{1}{2\sigma_X^2} + \frac{1}{2\sigma_Y^2}} - \frac{b^2}{2\sigma_Y^2}\right) \\
&= \frac{1}{\sqrt{2\pi}\sqrt{\sigma_X^2 + \sigma_Y^2}} \exp\left(\frac{b^2(\sigma_X^2/\sigma_Y^2)}{2(\sigma_X^2 + \sigma_Y^2)} - \frac{b^2}{2\sigma_Y^2}\right) \\
&= \frac{1}{\sqrt{2\pi}\sqrt{\sigma_X^2 + \sigma_Y^2}} \exp\left(\frac{b^2}{2(\sigma_X^2 + \sigma_Y^2)}\right)
\end{aligned}$$

where (i) is because the pdf of Gaussian integrates to 1. Hence $X + Y \sim N(0,\, \sigma_X^2 + \sigma_Y^2)$. In general, when $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$, we have

$$X + Y \sim N(\mu_X + \mu_Y,\, \sigma_X^2 + \sigma_Y^2).$$

(e) (4 pts) Answer: Part (d) says that the sum of independent Gaussian random variables also follows the Gaussian distribution. Central limit theorem says that the sum of i.i.d. random variables tends to Gaussian (under proper normalization), even if the distribution of the individual random variables is not Gaussian. Therefore Gaussian distribution is an "attractor" for i.i.d. sum.

Remark (not needed for full score): It is possible to show that the sum of two independent Gaussian random variables is also Gaussian using central limit theorem. Let $U_1, U_2, \ldots$ be i.i.d. with mean 0 and variance σ_1^2. Let $V_1, V_2, \ldots$ be i.i.d. with mean 0 and variance σ_2^2 independent of $U_1, U_2, \ldots$. By central limit theorem, $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_i$ tends to the distribution

$N(0, \sigma_1^2)$, and $\frac{1}{\sqrt{n}} \sum_{i=1}^n V_i$ tends to the distribution $N(0, \sigma_2^2)$, and hence the limiting distribution of $\frac{1}{\sqrt{n}} \sum_{i=1}^n U_i + \frac{1}{\sqrt{n}} \sum_{i=1}^n V_i$ should be the the same as that of the sum of two independent Gaussian random variables, where one follows $N(0, \sigma_1^2)$ and the other follows $N(0, \sigma_2^2)$. Again by central limit theorem, we also know that $\frac{1}{\sqrt{n}} \sum_{i=1}^n (U_i + V_i)$ tends to the distribution $N(0, \sigma_1^2 + \sigma_2^2)$. Hence we obtain the same result as part (d).

4. **(20 pts) Estimating mean and variance of Gaussian**

(a) (5 pts) Answer: Consider the log-likelihood

$$\begin{aligned}
\ln f(x_1, \ldots, x_n; \mu) &= \sum_{i=1}^n \ln \left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{(x_i - \mu)^2}{2\sigma^2} \right) \right) \\
&= -n \ln \left(\sqrt{2\pi\sigma^2} \right) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2.
\end{aligned}$$

Taking derivative,

$$\begin{aligned}
\frac{d}{d\mu} \ln f(x_1, \ldots, x_n; \mu) &= \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) \\
&= \frac{1}{\sigma^2} \left(\sum_{i=1}^n x_i - n\mu \right).
\end{aligned}$$

Hence log-likelihood is maximized at

$$\mu^* = \frac{1}{n} \sum_{i=1}^n x_i.$$

The maximum likelihood estimate is

$$\hat{\mu}_{\text{ML}} = \frac{1}{n} \sum_{i=1}^n X_i.$$

Since $\mathbb{E}[\hat{\mu}_{\text{ML}}] = \frac{1}{n} \sum_i \mathbb{E}[X_i] = \mu$, the estimator is unbiased. The mean square error

$$\begin{aligned}
\mathbb{E}[(\hat{\mu}_{\text{ML}} - \mu)^2] &= \operatorname{Var}(\hat{\mu}_{\text{ML}}) \\
&= \operatorname{Var} \left(\frac{1}{n} \sum_{i=1}^n X_i \right) \\
&= \frac{1}{n^2} \sum_{i=1}^n \operatorname{Var}(X_i) \\
&= \frac{\sigma^2}{n}
\end{aligned}$$

is inversely proportional to n.

(b) (7 pts) Answer: Consider the log-likelihood

$$\ln f(x_1, \ldots, x_n; \sigma^2) = -n \ln \left(\sqrt{2\pi\sigma^2} \right) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2.$$

Taking derivative,

$$\frac{d}{d(\sigma^2)} \ln f(x_1, \ldots, x_n; \sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (x_i - \mu)^2$$
$$= \frac{1}{2\sigma^4} \left(\sum_{i=1}^{n} (x_i - \mu)^2 - n\sigma^2 \right).$$

Hence log-likelihood is maximized at

$$(\sigma^*)^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2.$$

The maximum likelihood estimate is

$$\hat{\sigma}^2_{\mathrm{ML}} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2.$$

Since $\mathbb{E}[\hat{\sigma}^2_{\mathrm{ML}}] = \frac{1}{n} \sum_i \mathbb{E}[(X_i - \mu)^2] = \sigma^2$, the estimator is unbiased. The mean square error

$$\begin{aligned}
\mathbb{E}[(\hat{\sigma}^2_{\mathrm{ML}} - \sigma^2)^2] &= \mathrm{Var}\left(\hat{\sigma}^2_{\mathrm{ML}}\right) \\
&= \mathrm{Var}\left(\frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2 \right) \\
&= \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{Var}\left((X_i - \mu)^2\right) \\
&= \frac{1}{n^2} \sum_{i=1}^{n} \mathbb{E}\left[\left((X_i - \mu)^2 - \sigma^2 \right)^2 \right] \\
&= \frac{\sigma^4}{n^2} \sum_{i=1}^{n} \mathbb{E}\left[\left(\left(\frac{X_i - \mu}{\sigma} \right)^2 - 1 \right)^2 \right] \\
&= \frac{\sigma^4}{n^2} \sum_{i=1}^{n} \mathbb{E}\left[\left(\frac{X_i - \mu}{\sigma} \right)^4 - 2 \left(\frac{X_i - \mu}{\sigma} \right)^2 + 1 \right] \\
&= \frac{\sigma^4}{n^2} \sum_{i=1}^{n} (3 - 2 + 1) \\
&= \frac{2\sigma^4}{n}
\end{aligned}$$

is inversely proportional to n.

(c) (8 pts) Answer: Consider the log-likelihood

$$\ln f(x_1, \ldots, x_n; \mu, \sigma^2) = -n \ln\left(\sqrt{2\pi\sigma^2}\right) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2.$$

Taking (partial) derivative,

$$\frac{\partial}{\partial\mu} \ln f(x_1, \ldots, x_n; \mu, \sigma^2) = \frac{1}{\sigma^2} \left(\sum_{i=1}^{n} x_i - n\mu \right).$$

$$\frac{\partial}{\partial(\sigma^2)} \ln f(x_1,\ldots,x_n;\mu,\sigma^2) = \frac{1}{2\sigma^4}\left(\sum_{i=1}^{n}(x_i-\mu)^2 - n\sigma^2\right).$$

Setting both to 0, we obtain the maximum likelihood estimates

$$\hat{\mu}_{\mathrm{ML}} = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

$$\hat{\sigma}^2_{\mathrm{ML}} = \frac{1}{n}\sum_{i=1}^{n}(X_i-\hat{\mu}_{\mathrm{ML}})^2.$$

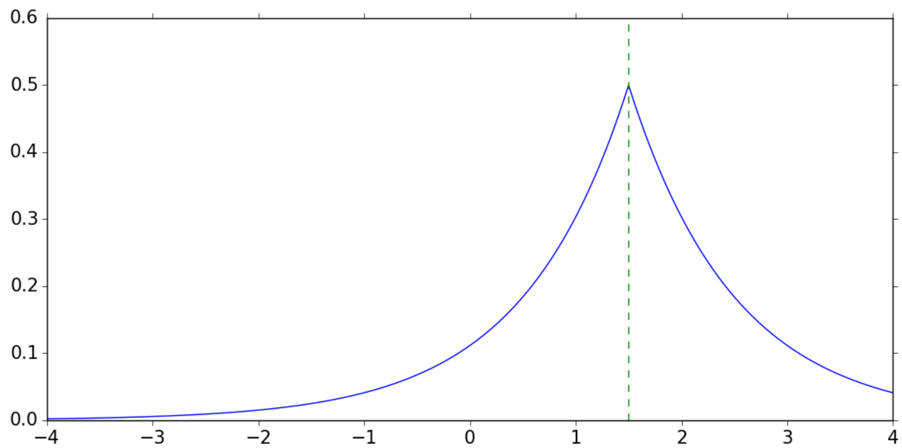
Since $\mathbb{E}[\hat{\mu}_{\mathrm{ML}}] = \frac{1}{n}\sum_i \mathbb{E}[X_i] = \mu$, $\hat{\mu}_{\mathrm{ML}}$ is unbiased. However,

$$\begin{aligned}
\mathbb{E}\left[\hat{\sigma}^2_{\mathrm{ML}}\right] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\hat{\mu}_{\mathrm{ML}})^2\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[(X_i - \frac{1}{n}\sum_{j=1}^{n}X_j)^2\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\left(\frac{n-1}{n}(X_i-\mu) - \frac{1}{n}\sum_{j\neq i}(X_j-\mu)\right)^2\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\mathbb{E}\left[\left(\frac{n-1}{n}(X_i-\mu)\right)^2\right] + \sum_{j\neq i}\mathbb{E}\left[\left(\frac{1}{n}(X_j-\mu)\right)^2\right]\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\left(\frac{n-1}{n}\right)^2\sigma^2 + \frac{n-1}{n^2}\sigma^2\right) \\
&= \left(\frac{n-1}{n}\right)^2\sigma^2 + \frac{n-1}{n^2}\sigma^2 \\
&= \frac{n-1}{n}\sigma^2.
\end{aligned}$$

Hence $\hat{\sigma}^2_{\mathrm{ML}}$ is a biased estimator.
Remark: Some people prefer using the unbiased estimator $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\hat{\mu}_{\mathrm{ML}})^2$ instead of the MLE estimator $\hat{\sigma}^2_{\mathrm{ML}}$.

5. **(20 pts) Maximum Likelihood of Laplace**

(a) (5 pts) Answer: The parameter θ is the center of the distribution (which is also the mean and median). The following is the plot of the pdf when $\theta = 1.5$ (the dotted line):



(b) (8 pts) Answer: Consider the log-likelihood

$$\begin{aligned}\ln f(x_1,\ldots,x_n;\theta) &= \sum_{i=1}^{n} \ln f(x_i;\theta)\\ &= \sum_{i=1}^{n} \ln\left(\frac{1}{2}\exp\left(-|x_i-\theta|\right)\right)\\ &= -n\ln 2 - \sum_{i=1}^{n}|x_i-\theta|.\end{aligned}$$

When $x_1 = 0$, $x_2 = 1$, $x_3 = 3$,

$$\sum_{i=1}^{n}|x_i-\theta| = \begin{cases} 4-3\theta & \text{if } \theta < 0\\ 4-\theta & \text{if } 0 \le \theta < 1\\ 2+\theta & \text{if } 1 \le \theta < 3\\ 3\theta-4 & \text{if } \theta \ge 3\end{cases}$$

Hence $\sum_{i=1}^{n}|x_i-\theta|$ is minimized at $\theta = 1$. The MLE is $\hat{\theta} = 1$.

(c) (7 pts) Answer: Consider the log-likelihood

$$\ln f(x_1,\ldots,x_n;\theta) = -n\ln 2 - \sum_{i=1}^{n}|x_i-\theta|.$$

Assume $x_1 \le \cdots \le x_n$. We would like to minimize $\sum_{i=1}^{n}|x_i-\theta|$. If $\theta > x_{\lceil (n+1)/2\rceil}$ (i.e., there are more than half of x_i's on the left of θ), then θ cannot be the minimizer of $\sum_{i=1}^{n}|x_i-\theta|$ since we can reduce θ to reduce its distance to more than half of x_i's, while increasing its distance to less than half of x_i's, and hence reducing the total distance. Similarly if $\theta < x_{\lceil n/2\rceil}$ then θ cannot be the minimizer. Hence $\sum_{i=1}^{n}|x_i-\theta|$ is minimized when

$$x_{\lceil n/2\rceil} \le \theta \le x_{\lceil (n+1)/2\rceil},$$

i.e., θ is the median of $x_1,\ldots,x_n$ (if n is even then any $x_{n/2} \le \theta \le x_{n/2+1}$ is a minimizer). Hence the MLE $\hat{\theta}_{\text{ML}}$ is the median of $x_1,\ldots,x_n$.