Massachusetts Institute of Technology

Department of Electrical Engineering & Computer Science

6.041/6.431: Probabilistic Systems Analysis

(Fall 2009)

Problem Set 11: Solutions
Due: Never

1. (a) Let C denote the coin that Bob received, so that $C = 1$ if Bob received first coin, and $C = 2$ if Bob received the second coin. Then $\mathbf{P}(C = 1) = p$ and $\mathbf{P}(C = 2) = 1 - p$. Given C, the number of heads Y in 3 independent tosses is a Binomial random variable. We can find the probability that Bob received the first coin given that he observed k heads using Bayes rule.

$$\begin{aligned}
\mathbf{P}(C = 1 \mid Y = k) &= \frac{\mathbf{P}(Y = k \mid C = 1) \cdot \mathbf{P}(C = 1)}{P(Y = k \mid C = 1) \cdot \mathbf{P}(C = 1) + \mathbf{P}(Y = k \mid C = 2) \cdot \mathbf{P}(C = 2)} \\
&= \frac{\binom{3}{k} \cdot (1/3)^k (2/3)^{3-k} p}{\binom{3}{k} \cdot (1/3)^k (2/3)^{3-k} \cdot p + \binom{3}{k} \cdot (2/3)^k (1/3)^{3-k} \cdot (1-p)} \\
&= \frac{2^{3-k} p}{2^{3-k} p + 2^k (1-p)} = \frac{1}{1 + \frac{1-p}{p} 2^{2k-3}}
\end{aligned}$$

(b) We want to find k so that the following inequality holds.

$$\begin{aligned}
\mathbf{P}(C = 1 \mid Y = k) &> p \\
\frac{2^{3-k} p}{2^{3-k} p + 2^k (1-p)} &> p
\end{aligned}$$

Note that if $p = 0$ or $p = 1$, there is no value of k that satisfies the inequality. We now solve it for $0 < p < 1$:

$$\begin{aligned}
\frac{2^{3-k}}{2^{3-k} p + 2^k (1-p)} &> 1 \\
2^{3-k} &> 2^{3-k} p + 2^k (1-p) \\
2^{3-k}(1-p) &> 2^k (1-p) \\
2^{3-k} &> 2^k \\
2k &< 3 \\
k &< 3/2
\end{aligned}$$

For $0 < p < 1$, $k = 0$ or $k = 1$ the probability that Alice sent the first coin increases. The inequality does not depend on p, and so does not change when p increases. Intuitively, this makes sense: lower values of k increase Bob's belief he got the coin with lower probability of heads.

(c) Given that Bob observes k heads, Bob must decide on whether the first or second coin was used. To minimize the error, he should decide it is the first coin when $\mathbf{P}(C = 1 \mid Y = k) \geq \mathbf{P}(C = 2 \mid Y = k)$. Thus, we have the decision rule given by

$$\begin{aligned}
\mathbf{P}(C=1 \mid Y=k) &\geq \mathbf{P}(C=2 \mid Y=k) \\
\frac{2^{3-k}p}{2^{3-k}p+2^k(1-p)} &\geq \frac{2^k(1-p)}{2^{3-k}p+2^k(1-p)} \\
2^{3-k}p &\geq 2^k(1-p) \\
2^{2k-3} &\leq \frac{p}{1-p} \\
k &\leq \frac{3}{2}+\frac{1}{2}\log_2\frac{p}{1-p}
\end{aligned}$$

(d) i. If $p = 2/3$, the threshold in the rule above is equal to $\frac{3+\log_2 2}{2} = 2$. Therefore, Bob will decide that he received the first coin when he observes 0, 1 or 2 heads, and will decide that he received the second coin when he observes 3 heads.
We find he probability of a correct decision using the total probability law:

$$\begin{aligned}
\mathbf{P}(\text{Correct}) &= \mathbf{P}(\text{Correct} \mid C=1)\cdot p + \mathbf{P}(\text{Correct} \mid C=2)\cdot(1-p) \\
&= \mathbf{P}(Y<3 \mid C=1)\cdot p + \mathbf{P}(Y=3 \mid C=2)\cdot(1-p) \\
&= (1-\mathbf{P}(Y=3 \mid C=1))\cdot p + \mathbf{P}(Y=3 \mid C=2)\cdot(1-p) \\
&= (1-(1/3)^3)(2/3)+(2/3)^3(1/3) = 20/27 \approx .741
\end{aligned}$$

ii. In absence of any data, all Bob can do is decide he received the first coin with some probability q. Note that this rule includes the deterministic decisions that he received either the first coin ($q = 1$) or the second coin ($q = 0$).
In this case, the probability of correct decision is equal to

$$\begin{aligned}
\mathbf{P}(\text{Correct}) &= \mathbf{P}(\text{Correct} \mid C=1)\cdot p + \mathbf{P}(\text{Correct} \mid C=2)\cdot(1-p) \\
&= qp+(1-q)(1-p) = 1-p+q(2p-1) = \frac{1+q}{3}
\end{aligned}$$

Clearly, the probability of the correct decision is maximized (or the probability of error is minimized) when $q = 1$, i.e., when Bob deterministically decides he received the first coin. In this case, $\mathbf{P}(\text{Correct}) = 2/3 \approx .667$. Observing 3 coin tosses increases the probability of the correct decision by $2/27 \approx .074$.

(e) If p is increased, the threshold in the decision rule in part (c) goes up, i.e., the range of values of k for which Bob decides he received the first coin can only go up.

(f) Bob will never decide he received the first coin if the threshold in the rule above is below zero:

$$\begin{aligned}
\frac{3}{2}+\frac{1}{2}\log_2\frac{p}{1-p} &< 0 \\
\log_2\frac{p}{1-p} &< -3 \\
\frac{p}{1-p} &< \frac{1}{8} \\
p &< \frac{1}{9}
\end{aligned}$$

If $p < 1/9$, the prior probability of receiving the first coin is so low that no amount of evidence from 3 tosses of the coin will make Bob decide he received the first coin.

(g) Bob will always decide he received the first coin if the threshold in the rule above is equal to or above 3:

$$\begin{aligned} \frac{3}{2} + \frac{1}{2}\log_2 \frac{p}{1-p} &\geq 3 \\ \log_2 \frac{p}{1-p} &\geq 3 \\ \frac{p}{1-p} &\geq 8 \\ p &\geq \frac{8}{9} \end{aligned}$$

If $p \geq 8/9$, the prior probability of receiving the first coin is so high that no amount of evidence from 3 tosses of the coin will make Bob decide he received the second coin.

2. (a) To find the MAP estimate, we need to find the value x that maximizes the conditional density $f_{X|Y}(x \mid y)$ by taking its derivative and setting it to 0.

$$\begin{aligned} f_{X|Y}(x \mid y) &= \frac{p_{Y|X}(y \mid x) \cdot f_X(x)}{p_Y(y)} \\ &= \frac{e^{-x}x^y}{y!} \cdot \mu e^{-\mu x} \cdot \frac{1}{p_Y(y)} \\ &= \frac{\mu}{y! p_Y(y)} \cdot e^{-(\mu+1)x} x^y \end{aligned}$$

$$\begin{aligned} \frac{d}{dx} f_{X|Y}(x \mid y) &= \frac{d}{dx}\left(\frac{\mu}{y! p_Y(y)} \cdot e^{-(\mu+1)x} x^y\right) \\ &= \frac{\mu}{y! p_Y(y)} \, x^{y-1} \, e^{-(\mu+1)x} \; (y - x(\mu+1)) \end{aligned}$$

Since the only factor that depends on x which can take on the value 0 is $(y - x(\mu + 1))$, the maximum is achieved at

$$\hat{x}_{\text{MAP}}(y) = \frac{y}{1+\mu}$$

It is easy to check that this value is indeed maximum (the first derivative changes from positive to negative at this value).

(b) i. To show the given identity, we need to use Bayes' rule. We first compute the denominator, $p_Y(y)$

$$\begin{aligned} p_Y(y) &= \int_0^\infty \frac{e^{-x}x^y}{y!} \mu e^{-\mu x} \, dx \\ &= \frac{\mu}{y! \, (1+\mu)^{y+1}} \int_0^\infty (1+\mu)^{y+1} x^y e^{-(1+\mu)x} \, dx \\ &= \frac{\mu}{(1+\mu)^{y+1}} \end{aligned}$$

Then, we can substitute into the equation we had derived in part (a)

$$\begin{aligned} f_{X|Y}(x \mid y) &= \frac{\mu}{y! p_Y(y)} x^y e^{-(\mu+1)x} \\ &= \frac{\mu}{y!} \frac{(1+\mu)^{y+1}}{\mu} x^y e^{-(\mu+1)x} \\ &= \frac{(1+\mu)^{y+1}}{y!} x^y e^{-(\mu+1)x} \end{aligned}$$

Thus, $\lambda = 1 + \mu$.

ii. We first manipulate $x f_{X|Y}(x \mid y)$:

$$\begin{aligned} x f_{X|Y}(x \mid y) &= \frac{(1+\mu)^{y+1}}{y!} x^{y+1} e^{-(\mu+1)x} \\ &= \frac{y+1}{1+\mu} \frac{(1+\mu)^{y+2}}{(y+1)!} x^{y+1} e^{-(\mu+1)x} \\ &= \frac{y+1}{1+\mu} f_{X|Y}(x \mid y+1) \end{aligned}$$

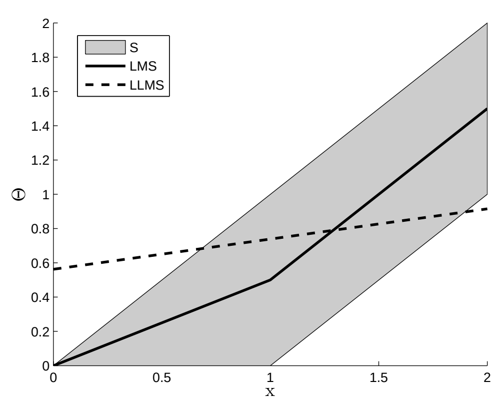
Now we can find the conditional expectation estimator:

$$\begin{aligned} \hat{x}_{\mathrm{CE}}(y) &= \mathbf{E}[X|Y=y] = \int_0^\infty x f_{X|Y}(x \mid y)\, dx \\ &= \int_0^\infty \frac{y+1}{1+\mu} f_{X|Y}(x \mid y+1)\, dx = \frac{y+1}{1+\mu} \end{aligned}$$

(c) The conditional expectation estimator is always higher than the MAP estimator by $\frac{1}{1+\mu}$.

3. (a)-(e) See online solutions of problem 16 on page 449 of the text.

(f) The sketch required is given below:



4. (a) To find the normalization constant c we integrate the joint PDF:

$$\int_0^1 \int_0^1 f_{X,Y}(x,y)\,dy\,dx = c\int_0^1 \int_0^1 xy\,dy\,dx = c\int_0^1 1/2x\,dx = c/4.$$

Therefore, $c = 4$.

(b) To construct the conditional expectation estimator, we need to find the conditional probability density.

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{4xy}{\int_0^1 4xy\,dx} = \frac{4xy}{2y} = 2x, \quad x \in (0,1]$$

Thus

$$\hat{x}_{\mathrm{CE}}(y) = \mathbf{E}[X|Y=y] = \int_0^1 x \cdot 2x\,dx = 2/3.$$

(c) We first note that the conditional probability does not depend on y. Therefore, X and Y are independent, and whether or not we observe $Y = y$ does not affect the estimate in part (b).

Another way to see this is to consider that if we do not observe y, we can compute the marginal $f_X(x) = \int_0^1 4xydy = 2x$ which is equal to the conditional density, and will therefore produce the same estimate.

(d) Since X and Y are independent, no estimator can make use of the observed value of Y to estimate X. The MAP estimator for X is equal to 1, regardless of what value y we observe, since the conditional (and the marginal) density is maximized at 1.

5. (a) X is a binomial random variable with parameters $n = 3$ and given the probability p that a single bit is flipped in a transmission over the noisy channel:

$$p_X(k;p) = \begin{cases} \binom{3}{k} p^k (1-p)^{3-k}, & k = 0,1,2,3 \\ 0 & \text{o.w.} \end{cases}$$

(b) To derive the ML estimator for p based on $X_1, \ldots, X_n$, the numbers of bits flipped in the first n three-bit messages, we need to find the value of p that maximizes the likelihood function:

$$\hat{p}_n = \arg\max_p \, p_{X_1,\ldots,X_n}(k_1, k_2, \ldots, k_n; p)$$

Since the X_i's are independent, the likelihood function simplifies to:

$$p_{X_1,\ldots,X_n}(k_1, k_2, \ldots, k_n; p) = \Pi_{i=1}^n p_{X_i}(k_i;p) = \Pi_{i=1}^n \binom{3}{k_i} p^{k_i}(1-p)^{3-k_i}$$

The log-likelihood function is given by

$$\log(p_{X_1,\ldots,X_n}(k_1, k_2, \ldots, k_n; p)) = \sum_{i=1}^n \left(k_i \log(p) + (3-k_i)\log(1-p) + \log\binom{3}{k_i}\right)$$

We then maximize the log-likelihood function with respect to p:

$$\begin{aligned} \frac{1}{p}\left(\sum_{i=1}^{n} k_i\right) - \frac{1}{1-p}\left(\sum_{i=1}^{n}(3-k_i)\right) &= 0 \\ \left(3n - \sum_{i=1}^{n} k_i\right) p &= \left(\sum_{i=1}^{n} k_i\right)(1-p) \\ \hat{p}_n = \frac{1}{3n}\sum_{i=1}^{n} k_i \end{aligned}$$

This yields the ML estimator:

$$\hat{P}_n = \frac{1}{3n}\sum_{i=1}^{n} X_i$$

(c) The estimator is unbiased since:

$$\begin{aligned} \mathbf{E}_p[\hat{P}_n] &= \frac{1}{3n}\sum_{i=1}^{n}\mathbf{E}_p[X_i] \\ &= \frac{1}{3n}\sum_{i=1}^{n} 3p \\ &= p \end{aligned}$$

(d) By the weak law of large numbers, $\frac{1}{n}\sum_{i=1}^{n} X_i$ converges in probability to $\mathbf{E}_p[X_i] = 3p$, and therefore $\hat{P}_n = \frac{1}{3n}\sum_{i=1}^{n} X_i$ converges in probability to p. Thus $\hat{P}_n$ is consistent.

6. (a) Normalization of the distribution requires:

$$1 = \sum_{k=0}^{\infty} p_K(k;\theta) = \sum_{k=0}^{\infty}\frac{e^{-\frac{k}{\theta}}}{Z(\theta)} = \frac{1}{Z(\theta)}\sum_{k=0}^{\infty} e^{-\frac{k}{\theta}} = \frac{1}{Z(\theta)\cdot(1-e^{-\frac{1}{\theta}})},$$

so $Z(\theta) = \frac{1}{1-e^{-\frac{1}{\theta}}}$.

(b) Rewriting $p_K(k;\theta)$ as:

$$p_K(k;\theta) = \left(e^{-\frac{1}{\theta}}\right)^k\left(1-e^{-\frac{1}{\theta}}\right),$$

the probability distribution for the photon number is a geometric probability distribution with probability of success $p = 1-e^{-\frac{1}{\theta}}$, and it is shifted with 1 to the left since it starts with $k = 0$. Therefore the photon number expectation value is

$$\mu_K = \frac{1}{p} - 1 = \frac{1}{1-e^{-\frac{1}{\theta}}} - 1 = \frac{1}{e^{\frac{1}{\theta}} - 1}$$

and its variance is

$$\sigma_K^2 = \frac{1-p}{p^2} = \frac{e^{-\frac{1}{\theta}}}{(1-e^{-\frac{1}{\theta}})^2} = \mu_K^2 + \mu_K.$$

(c) The joint probability distribution for the k_i is

$$p_K(k_1, ..., k_n; \theta) = \frac{1}{Z(\theta)^n} \Pi_{i=1}^n e^{-k_i/\theta} = \frac{1}{Z(\theta)^n} e^{-\frac{1}{\theta}\sum_{i=1}^n k_i}.$$

The log likelihood is $-n \cdot \log Z(\theta) - 1/\theta \sum_{i=1}^n k_i$.
We find the maxima of the log likelihood by setting the derivative with respect to the parameter θ to zero:

$$\frac{d}{d\theta} \log p_K(k_1, ..., k_n; \theta) = -n \cdot \frac{e^{-\frac{1}{\theta}}}{\theta^2(1 - e^{-\frac{1}{\theta}})} + \frac{1}{\theta^2} \sum_{i=1}^n k_i = 0$$

or

$$\frac{1}{e^{\frac{1}{\theta}} - 1} = \frac{1}{n} \sum_{i=1}^n k_i = s_n.$$

For a hot body, $\theta \gg 1$ and $\frac{1}{e^{\frac{1}{\theta}} - 1} \approx \frac{1}{\theta}$, we obtain

$$\theta = \frac{1}{n} \sum_{i=1}^n k_i = s_n.$$

Thus the maximum likelihood estimator $\hat{\Theta}_n$ for the temperature is given in this limit by the sample mean of the photon number

$$\hat{\Theta}_n = \frac{1}{n} \sum_{i=1}^n K_i.$$

(d) According to the central limit theorem, the sample mean approaches for large n a Gaussian distribution with standard deviation our root mean square error

$$\sigma_{\hat{\Theta}_n} = \frac{\sigma_K}{\sqrt{n}}.$$

To allow only for 1% relative root mean square error in the temperature, we need $\frac{\sigma_K}{\sqrt{n}} < 0.01\mu_K$. With $\sigma_K^2 = \mu_K^2 + \mu_K$ it follows that

$$\sqrt{n} > \frac{\sigma_K}{0.01\mu_K} = 100 \frac{\sqrt{\mu_K^2 + \mu_K}}{\mu_K} = 100 \sqrt{1 + \frac{1}{\mu_K}}.$$

In general, for large temperatures, i.e. large mean photon numbers $\mu_K \gg 1$, we need about 10,000 samples.

(e) The 95% confidence interval for the temperature estimate for the situation in part (d), i.e.

$$\sigma_{\hat{\Theta}_n} = \frac{\sigma_K}{\sqrt{n}} = 0.01\mu_K,$$

is

$$[\hat{K} - 1.96\sigma_{\hat{K}}, \hat{K} + 1.96\sigma_{\hat{K}}] = [\hat{K} - 0.01\mu_K, \hat{K} + 0.01\mu_K].$$

7. (a) The variance of the estimator $\hat{\Delta}$ is the sum of the variance of the estimators $\hat{\Delta}_b$ and $\hat{\Delta}_b$:

$$\begin{aligned}
\mathrm{var}(\hat{\Delta}) &= \mathrm{var}(\hat{\Delta}_a) + \mathrm{var}(\hat{\Delta}_b) \\
&= \frac{\mathrm{var}(X_1^a + \cdots + X_{n_1}^a)}{n_1^2} + \frac{\mathrm{var}(X_1^b + \cdots + X_{n_2}^b)}{n_2^2} \\
&= \frac{\sigma_a^2}{n_1} + \frac{\sigma_b^2}{n_2}
\end{aligned}$$

We have:

$$P\left(\frac{|\hat{\Delta}_a + \hat{\Delta}_b - (\delta_a + \delta_b)|}{(\frac{\sigma_a^2}{n_1} + \frac{\sigma_b^2}{n_2})^{1/2}} \le 1.96\right) \quad \ge \quad .95$$

yielding a 95% confidence interval for $\delta_a + \delta_b$ of the form:

$$\left[\hat{\Delta}_a + \hat{\Delta}_b - 1.96\left(\frac{\sigma_a^2}{n_1} + \frac{\sigma_b^2}{n_2}\right)^{1/2}, \ \hat{\Delta}_a + \hat{\Delta}_b + 1.96\left(\frac{\sigma_a^2}{n_1} + \frac{\sigma_b^2}{n_2}\right)^{1/2}\right]$$

(b) A 97.5% confidence interval corresponds to an area of .0125 in each tail of the standard normal distribution. Using the normal CDF table, these tails correspond to the density at values greater than 2.24:

$$P\left(\frac{|\hat{\Delta}_a - \delta_a|}{\sigma_a/\sqrt{n_1}} > 2.24\right) = .025$$

$$P\left(\frac{|\hat{\Delta}_b - \delta_b|}{\sigma_b/\sqrt{n_1}} > 2.24\right) = .025$$

Let A be the event $\left\{\frac{|\hat{\Delta}_a - \delta_a|}{\sigma_a/\sqrt{n_1}} > 2.24\right\}$, and let B be the event $\left\{\frac{|\hat{\Delta}_b - \delta_b|}{\sigma_b/\sqrt{n_1}} > 2.24\right\}$. The union bound tells us that $P(A \cup B) \le P(A) + P(B)$. Using De Morgan's Law, $P(A \cup B) = 1 - P(A^c \cap B^c)$, which gives us $P(A^c \cap B^c) \ge 1 - (P(A) + P(B))$. Substituting our expressions for events A and B in this inequality yields:

$$\begin{aligned}
P\left(\frac{|\hat{\Delta}_a - \delta_a|}{\sigma_a/\sqrt{n_1}} \le 2.24, \frac{|\hat{\Delta}_b - \delta_b|}{\sigma_b/\sqrt{n_2}} \le 2.24\right) &\ge .95 \\
P\left(\hat{\Delta}_a + \hat{\Delta}_b - 2.24\left(\frac{\sigma_a}{\sqrt{n_1}} + \frac{\sigma_b}{\sqrt{n_2}}\right) \le \delta_a + \delta_b \le \hat{\Delta}_a + \hat{\Delta}_b + 2.24\left(\frac{\sigma_a}{\sqrt{n_1}} + \frac{\sigma_b}{\sqrt{n_2}}\right)\right) &\ge .95
\end{aligned}$$

where the second inequality follows since $P(X_1 \le \epsilon_1, X_2 \le \epsilon_2) \le P(X_1 + X_2 \le \epsilon_1 + \epsilon_2)$. Our confidence interval for $\delta_a + \delta_b$ in this case is:

$$\left[\hat{\Delta}_a + \hat{\Delta}_b - 2.24\left(\frac{\sigma_a}{\sqrt{n_1}} + \frac{\sigma_b}{\sqrt{n_2}}\right), \ \hat{\Delta}_a + \hat{\Delta}_b + 2.24\left(\frac{\sigma_a}{\sqrt{n_1}} + \frac{\sigma_b}{\sqrt{n_2}}\right)\right]$$

(c) We expect the bound with the lower variance estimate to be the tighter confidence interval. This corresponds to the direct estimate of the variance given in part (a). We can verify this by looking at the squared ratio of widths of the confidence intervals $\left(\frac{CI_{(a)}}{CI_{(b)}}\right)^2$:

$$\frac{4 \cdot 1.96^2 \left(\frac{\sigma_a^2}{n_1} + \frac{\sigma_b^2}{n_2}\right)}{4 \cdot 2.24^2 \left(\frac{\sigma_a^2}{n_1} + \frac{\sigma_b^2}{n_2} + 2\frac{\sigma_a \sigma_b}{\sqrt{n_1 n_2}}\right)}$$

(d) We'd use method (a) in order to have the narrowest width of the confidence interval. We can choose the best values of n_1 and n_2 by finding the minimum of the estimator variance (or equivalently, the standard deviation) by differentiating our equation for the variance, setting it equal to zero, and solving for the roots. This procedure is shown below:

$$\begin{aligned} \frac{d}{dn_1}\left(\frac{\sigma_a^2}{n_1} + \frac{\sigma_b^2}{n - n_1}\right) &= \frac{-\sigma_a^2}{n_1^2} + \frac{\sigma_b^2}{(n-n_1)^2} \\ &= \frac{\sigma_b^2 n_1^2 - \sigma_a^2 (n-n_1)^2}{n_1^2 (n-n_1)^2} \end{aligned}$$

Setting the derivative to zero yields:

$$\sigma_b^2 n_1^2 = \sigma_a^2 (n - n_1)^2$$

The variance is minimized when $n_1 = \frac{n}{\frac{\sigma_b}{\sigma_a} + 1}$

(e) If the X_i^a's and X_i^b's are Bernoulli, both the mean and variance depend on the respective parameters θ_a and θ_b. To find the confidence interval, we would either need to estimate the variance using another estimator (two example estimators are given on page 468 of the course text), or we could use the worst case estimate of the variance $\theta(1-\theta) = \frac{1}{4}$. To make the confidence interval as narrow as possible we could again use the method in part (a), and similarly solve for n_1 and n_2 as we did in part (d).

G1†. (a) i. We start by deriving the conditional CDF of Y given $X = x$, $F_{Y|X}(y \mid x)$. We have

$$\begin{aligned} F_{Y|X}(y \mid x) &= \mathbf{P}(Y \le y | X = x) \\ &= \mathbf{P}(X \cos W \le y | X = x) \\ &= \mathbf{P}(\cos W \le \frac{y}{x}). \end{aligned}$$

We note that cos is a one-to-one decreasing function over $[0, \pi/2]$; so for $0 \le y \le x$,

$$F_{Y|X}(y \mid x) = \mathbf{P}(W \ge \arccos\frac{y}{x}) = 1 - \frac{2}{\pi}\arccos\frac{y}{x}.$$

Differentiation yields

$$f_{Y|X}(y \mid x) = \frac{2}{\pi\sqrt{x^2 - y^2}}, \quad 0 \le y \le x.$$

ii. Since we have the conditional distribution of Y given X and the marginal distribution of X, it is straightforward to obtain the joint distribution. Thus,

$$f_{X,Y}(x,y) = f_{Y|X}(y \mid x) f_X(x) = \frac{2}{\pi l \sqrt{x^2 - y^2}}, \quad 0 \le x \le l, 0 \le y \le x.$$

So the joint PDF is non-zero over the triangular region $\{(x,y) \mid 0 \le x \le l, 0 \le y \le x\}$ or equivalently $\{(x,y) \mid 0 \le y \le l, y \le x \le l\}$. To obtain $f_Y(y)$, we integrate the joint PDF over x:

$$\begin{aligned} f_Y(y) &= \frac{2}{\pi l} \int_y^l \frac{1}{\sqrt{x^2 - y^2}} dx = \ln\left(x + \sqrt{x^2 - y^2}\right)\Big|_y^l \\ &= \frac{2}{\pi l} \ln\left(\frac{l + \sqrt{l^2 - y^2}}{y}\right), \quad 0 \le y \le l, \end{aligned}$$

where we have used the hint to evaluate the definite integral.

iii. Now

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{1}{\ln\left(\frac{l+\sqrt{l^2-y^2}}{y}\right)\sqrt{x^2 - y^2}}, \quad y \le x \le l.$$

So the least-squares estimate is given by

$$\begin{aligned} \mathbf{E}[X|Y=y] &= \int_{-\infty}^{\infty} x f_{X|Y}(x|y) dy = \frac{1}{\ln\left(\frac{l+\sqrt{l^2-y^2}}{y}\right)} \int_y^l \frac{x}{\sqrt{x^2 - y^2}} dx \\ &= \frac{\sqrt{x^2 - y^2}\Big|_y^l}{\ln\left(\frac{l+\sqrt{l^2-y^2}}{y}\right)} = \frac{\sqrt{l^2 - y^2}}{\ln\left(\frac{l+\sqrt{l^2-y^2}}{y}\right)}, \quad 0 \le y \le l. \end{aligned}$$

It is worth noting that $\lim_{y \to 0} \mathbf{E}[X \mid Y = y] = 0$ and that $\lim_{y \to l} \mathbf{E}[X \mid Y = y] = l$, which matches our intuition of the problem.

iv. See solution for (b) ii.

(b) i. The linear least-squares estimate is given by

$$g(Y) = \mathbf{E}[X] + \rho \frac{\sigma_X}{\sigma_Y}(Y - \mathbf{E}[Y]) = \mathbf{E}[X] + \frac{\text{cov}(X,Y)}{\sigma_Y^2}(Y - \mathbf{E}[Y]).$$

Since X is uniformly-distributed between 0 and l, it follows that $\mathbf{E}[X] = l/2$. We can obtain $\mathbf{E}[Y]$ and $\mathbf{E}[Y^2]$ by using the law of iterated expectations and keeping in mind that X and W are independent, as follows.

$$\begin{aligned} \mathbf{E}[Y] &= \mathbf{E}[\mathbf{E}[Y \mid X]] = \mathbf{E}[\mathbf{E}[X \cos W \mid X]] = \mathbf{E}[X \mathbf{E}[\cos W]] = \mathbf{E}[X]\mathbf{E}[\cos W] \\ &= \mathbf{E}[X] \cdot \frac{2}{\pi} \int_0^{\pi/2} \cos w dw = \frac{l}{2} \cdot \frac{2}{\pi} [\sin w]_0^{\pi/2} = \frac{2}{\pi}, \end{aligned}$$

and

$$\begin{aligned}\mathbf{E}[Y^2] &= \mathbf{E}[\mathbf{E}[Y^2 \mid X]] = \mathbf{E}[\mathbf{E}[X^2\cos^2 W \mid X]] = \mathbf{E}[X^2\mathbf{E}[\cos^2 W]] = \mathbf{E}[X^2]\mathbf{E}[\cos^2 W] \\ &= \frac{1}{l}\int_0^l x^2 dx \cdot \frac{2}{\pi}\int_0^{\pi/2} \cos^2 w dw = \frac{l^3}{3l}\cdot\frac{1}{\pi}\int_0^{\pi/2}(1+\cos 2w)dw = \frac{l^2}{3\pi}\cdot\frac{\pi}{2} = \frac{l^2}{6}.\end{aligned}$$

So

$$\text{var}(Y) = \frac{l^2}{6} - \frac{l^2}{\pi^2} = \frac{l^2(\pi^2-6)}{6\pi^2}.$$

Now,

$$\mathbf{E}[XY] = \mathbf{E}[X^2\cos W] = \mathbf{E}[X^2]\mathbf{E}[\cos W] = \frac{l^2}{3}\cdot\frac{2}{\pi} = \frac{2l^2}{3\pi}.$$

Hence

$$\text{cov}(X,Y) = \frac{2l^2}{3\pi} - \frac{l}{2}\cdot\frac{l}{\pi} = \frac{l^2}{\pi}\left(\frac{2}{3}-\frac{1}{2}\right) = \frac{l^2}{6\pi}.$$

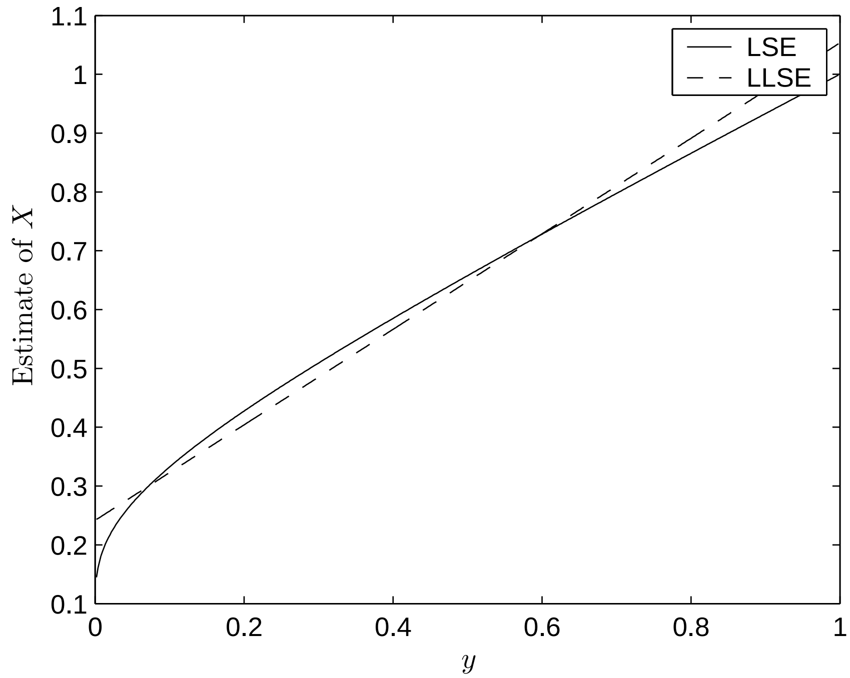
Therefore

$$g(Y) = \frac{l}{2} + \frac{l^2}{6\pi}\cdot\frac{6\pi^2}{l^2(\pi^2-6)}\left(Y - \frac{l}{\pi}\right) = \frac{l}{2} + \frac{\pi}{\pi^2-6}\left(Y-\frac{l}{\pi}\right).$$

The estimation error is given by

$$\begin{aligned}(1-\rho^2)\sigma_X^2 &= \left(\sigma_X^2 - \frac{\text{cov}^2(X,Y)}{\sigma_Y^2}\right) \\ &= \left(\frac{l^2}{12} - \frac{l^4}{36\pi^2}\cdot\frac{6\pi^2}{l^2(\pi^2-6)}\right) \\ &= \frac{l^2}{12}\left(1 - \frac{2}{\pi^2-6}\right) \\ &= \frac{l^2}{12}\cdot\frac{\pi^2-8}{\pi^2-6}.\end{aligned}$$

ii. The following plot shows the least-squares estimator (LSE) and linear least-squares estimator (LLSE) as a function of the observed projected length y for a needle of length $l = 1$.



(c) No, the intersection point is not uniform on the surface of the sphere. Consider fixing the elevation angle Ψ, and let the azimuth angle Θ vary. Since they are independent, Θ is still uniform between 0 and 2π for all values of Ψ.

If Ψ is small, then the resulting point is uniform on a circle of some radius, with a corresponding circumference. On the other hand, if Ψ is large, then the resulting point is uniform on a smaller circle.

Therefore the probability is not uniform on the sphere and in fact concentrated near the poles.

Our goal now is to make this intersection point uniformly distributed on the sphere.

Let us use (X, Y, Z) to denote the cartesian coordinate of the intersection point. Note: this has nothing to do with the random variables defined earlier in the problem. Also, for simplicity in notation, let us take Ψ as the angle from the z axis instead of from the xy-plane, so instead of taking values between $-\pi/2$ to $\pi/2$, it takes values between 0 to π.

Since the surface area of a sphere with radius r is $4\pi r^2$, the joint PDF (of the uniform distribution on the sphere) we are after has the value $\frac{1}{4\pi r^2}$, for $0 \leq x, y, z \leq r$ such that $x^2 + y^2 + z^2 = r^2$. So the joint CDF takes the form

$$\int\int\int \frac{1}{4\pi r^2} dxdydz.$$

We can transform the cartesian into spherical coordinates to get:

$$\begin{aligned} X &= R\sin(\Psi)\cos(\Theta) \\ Y &= R\sin(\Psi)\sin(\Theta) \\ Z &= R\cos(\Psi). \end{aligned}$$

In cartesian the change in volume was $dxdydz$; in spherical it is:

$$(dr)(rd\psi)(r\sin(\psi)d\theta).$$

(To get this recall that the arc length subtending an angle τ is $r\tau$.)
The CDF then takes the form:

$$\begin{aligned} & \int\int\int \frac{1}{4\pi r^2}(dr)(rd\psi)(r\sin(\psi)d\theta) \\ = & \int\int\int \frac{1}{4\pi r^2} r^2 \sin(\psi) dr d\psi d\theta \\ = & \int\int\int 1 \cdot \frac{\sin(\psi)}{2} \cdot \frac{1}{2\pi} dr d\psi d\theta. \end{aligned}$$

Therefore, one possibility is to have:

$$\begin{aligned} f_\Theta(\theta) &= \frac{1}{2\pi}, 0 \le \theta \le 2\pi \\ f_\Psi(\psi) &= \frac{\sin(\psi)}{2}, 0 \le \psi \le \pi \\ f_R(r) &= 1, 0 \le r \le 1. \end{aligned}$$

(Check that they integrate to 1.)