Probability and Computation: Problem sheet 3 Solutions

Question 1. *We are going to prove part of nicer version of the Chernoff Bounds. Prove the following inequalities*

i) For $0 < \delta < 1$,

$$\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}} \leq e^{-\delta^2/3}.$$

ii) For $0 < \delta < 1$,

$$\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}} \leq e^{-\delta^2/2}.$$

iii) Using the Chernoff-Bounds (Slides 8 and 22, Lecture 5) deduce the second part of the Nicer Chernoff Bounds in Slide 23 of Lecture 5.

Solution: i) By taking logarithm the problem is equivalent to prove that

$$\delta - (1+\delta)\log(1+\delta) + \delta^2/3 \leq 0$$

write $f(\delta) = 2\delta/3 - \log(1+\delta)$. We will check that $f(\delta) \geq 0$ for $\delta \in [0,1]$.

A straightforward computation shows that

- $f(0) = 0$,
- $f'(\delta) = 2\delta/3 - \log(1+\delta)$.
- $f'(0) = 0$ and $f'(1) = 2/3 - \log(2) < 0$

We deduce that $f'(\delta) \leq 0$ for $\delta \in [0,1]$. This holds because the function $\log(1+\delta)$ only intersects with the line $2x/3$ in exactly two points, one of them is 0 and the other has to be after 1 (otherwise $f'(1) \geq 0$). Finally,

$$f(\delta) = \int_0^{\delta} f'(x)dx - f(0) \leq \int_0^{\delta} 0dx - 0 = 0$$

ii) it is done the same way. Alternatively, use the Taylor expansion of $\log(1-\delta)$,

$$\log(1-\delta) = -\delta - \delta^2/2 - \delta^3/3 - \ldots \leq -\delta - \delta^2/2.$$

Question 2. *Chernoff Bounds for other random variables.*

i) Let X be a Poisson random variable of mean μ. Prove that

$$\mathbf{E}[\, e^{\lambda X} \,] = e^{\mu(e^{\lambda}-1)}$$

and deduce that for $t \geq \mu$ and for $\delta \in (0,1)$

$$\mathbf{P}[\, X \geq t \,] \leq e^{-\mu}\left(\frac{e\mu}{t}\right)^t \quad \textit{and} \quad \mathbf{P}[\, X \geq (1+\delta)\mu \,] \leq e^{-\delta^2\mu/3},$$

and the corresponding lower tails.

ii) Let X be a Normal random variable of mean μ and variance σ^2. Prove that

$$\mathbf{E}[\, e^{\lambda X} \,] = e^{\mu\lambda + \sigma^2\lambda^2/2},$$

and deduce that for $t > \mu$

$$\mathbf{P}[\, X \geq t \,] \leq \exp\left[\frac{-(t-\mu)^2}{2\sigma^2}\right].$$

Find the corresponding lower tail.

Solution: i) Recall that $\mathbf{P}[\, X = k \,] = \frac{\mu^k e^{-\mu}}{k!}$. Then

$$\mathbf{E}[\, e^{\lambda X} \,] = \sum_{i=0}^{\infty} e^{\lambda k} \frac{\mu^k e^{-\mu}}{k!} = \sum_{i=0}^{\infty} \frac{(e^{\lambda}\mu)^k e^{-\mu}}{k!} = \frac{e^{-\mu}}{e^{-\mu e^{\lambda}}} \sum_{i=0}^{\infty} \frac{(e^{\lambda}\mu)^k e^{-\mu e^{\lambda}}}{k!} \tag{1}$$

Note that $\frac{(e^{\lambda}\mu)^k e^{-\mu e^{\lambda}}}{k!}$ is the probability that a Poisson random variable of mean μe^{λ} takes value k, therefore the summation on the RHS of (1) is 1. We conclude

$$\mathbf{E}[\, e^{\lambda X} \,] = \frac{e^{-\mu}}{e^{-\mu e^{\lambda}}} = e^{\mu(e^{\lambda}-1)}$$

The computation of $\mathbf{P}[\, X \geq t \,]$ is exactly as in Lecture 5 slide 20.

ii) The density function of a normal random variable with mean μ and variance σ is given by $\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$. Then

$$\mathbf{E}[\, e^{\lambda X} \,] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} \exp(\lambda x) \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} \exp\left(-\frac{(x-\mu)^2 - 2\sigma^2\lambda x}{2\sigma^2}\right) dx$$

We analyse the expression inside the exponential

$$(x-\mu)^2 - 2\sigma^2\lambda x = x^2 - 2(\mu + \lambda\sigma^2)x + \mu^2 = (x - (\mu + \lambda\sigma^2))^2 + \mu^2 - (\mu + \lambda\sigma^2)^2$$

Therefore

$$\mathbf{E}[\, e^{\lambda X} \,] = \exp\left(-\frac{\mu^2 - (\mu + \lambda\sigma^2)^2}{2\sigma^2}\right) \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} \exp\left(-\frac{(x - (\mu + \lambda\sigma^2))^2}{2\sigma^2}\right) dx$$

Note that $\exp\left(-\frac{(x-(\mu+\lambda\sigma^2))^2}{2\sigma^2}\right)$ is the density function of a normal random variable with mean $\mu + \lambda\sigma^2$ and variance σ^2. Therefore we have

$$\mathbf{E}[\, e^{\lambda X} \,] = \exp\left(-\frac{\mu^2 - (\mu + \lambda\sigma^2)^2}{2\sigma^2}\right) = \exp\left(\mu\lambda + \frac{\sigma^2\lambda^2}{2}\right)$$

To compute $\mathbf{P}[\, X \geq t \,]$ use the recipe (Lecture 5, Slide 21).

Question 3. *Show properties 2-6 of slide 19 of Lecture 6.*

Solution: Property 4: Let a be a possible value of X. Since X is independent of Y we have $\mathbf{P}[\, Y = y | X = a \,] = \mathbf{P}[\, Y = y \,]$ then

$$\mathbf{E}[\, Y | X = a \,] = \sum_{y} y \mathbf{P}[\, Y = y | X = a \,] \overset{indep}{=} \sum_{y} y \mathbf{P}[\, Y = y \,] = \mathbf{E}[\, Y \,],$$

deducing that $\mathbf{E}[\, Y | X \,] = \mathbf{E}[\, Y \,]$.

Property 5: Let a be a possible value of X, then

$$\mathbf{E}[\,YZ|X=a\,] = \mathbf{E}[\,F(X)Z|X=a\,] = \sum_x \sum_z F(x)z\mathbf{P}[\,X=x, Z=z|X=a\,]$$
$$= F(a)\sum_z z\mathbf{P}[\,Z=z|X=a\,] = F(a)\mathbf{E}[\,Z|X=a\,],$$

deducing that $\mathbf{E}[\,YZ|X\,] = Y\mathbf{E}[\,Z|X\,]$.

Question 4. *Let $X_1,\ldots,X_n$ be independent discrete random variables and let $Z = f(X_1,\ldots,X_n)$ for some function f. Prove that*

$$\mathbf{E}[\,Z|X_1,\ldots,X_i\,] = \sum_{x_i,x_{i+1},\ldots,x_n} f(X_1,\ldots,X_i,x_i,\ldots,x_n)\mathbf{P}[\,X_i=x_i,\ldots,X_n=x_n\,]$$

Solution: Let $(a_1,\ldots,a_n)$ be a possible value of $(X_1,\ldots,X_n)$, then

$$\mathbf{E}[\,Z|X_1=a_1,\ldots,X_n=a_n\,] = \sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} f(x_1,\ldots,x_n)\mathbf{P}[\,X_1=x_1,\ldots,X_n=x_n|X_1=a_1,\ldots,X_i=a_i\,]$$
$$= \sum_{x_{i+1}}\cdots\sum_{x_n} f(a_1,\ldots,a_i,x_{i+1},\ldots,x_n)\mathbf{P}[\,X_{i+1}=x_{i+1},\ldots,X_n=x_n\,].$$

Question 5. *Conditional Variance. Define the conditional variance of Y given X as*

$$\mathbf{Var}\,[\,Y|X\,] = \mathbf{E}\big[\,(Y-\mathbf{E}[\,Y|X\,])^2|X\,\big].$$

1. *Prove that $\mathbf{Var}\,[\,Y\,] = \mathbf{E}[\,\mathbf{Var}\,[\,Y|X\,]\,] + \mathbf{Var}\,[\,\mathbf{E}[\,Y|X\,]\,]$*
2. *Consider n bins and a random number M of balls, where $\mathbf{E}[\,M\,]=\mu$ and $\mathbf{Var}\,[\,M\,]=\sigma^2$. Compute the variance of the number of balls that are assigned to the first bin.*

Solution: Remember that $\mathbf{Var}\,[\,Y\,] = \mathbf{E}\big[\,(Y-\mathbf{E}[\,Y\,])^2\,\big]$, and, equivalently, $\mathbf{Var}\,[\,Y\,] = \mathbf{E}\big[\,Y^2\,\big] - \mathbf{E}[\,Y\,]^2$

By definition we have that

$$\mathbf{E}[\,\mathbf{Var}\,[\,Y|X\,]\,] = \mathbf{E}\big[\,\mathbf{E}\big[\,(Y-\mathbf{E}[\,Y|X\,])^2|X\,\big]\,\big] \overset{p1}{=} \mathbf{E}\big[\,(Y-\mathbf{E}[\,Y|X\,])^2)\,\big] = \mathbf{E}\Big[\,Y^2 - 2Y\mathbf{E}[\,Y|X\,] + \mathbf{E}[\,Y|X\,]^2\,\Big]$$

($p1$ refers to the properties of Lecture 6, slide 19). By $p1$ we get

$$\mathbf{E}[\,Y\mathbf{E}[\,Y|X\,]\,] \overset{p1}{=} \mathbf{E}[\,\mathbf{E}[\,Y\mathbf{E}[\,Y|X\,]\,|X\,]\,] \overset{p5}{=} \mathbf{E}\Big[\,\mathbf{E}[\,Y|X\,]^2\,\Big]$$

by linearity of conditional expectation (p3) we get

$$\mathbf{E}[\,\mathbf{Var}\,[\,Y|X\,]\,] = \mathbf{E}\big[\,Y^2\,\big] - \mathbf{E}\Big[\,\mathbf{E}[\,Y|X\,]^2\,\Big] \tag{2}$$

Also, note that $\mathbf{E}[\,\mathbf{E}[\,Y|X\,]\,] = \mathbf{E}[\,Y\,]$ then

$$\mathbf{Var}\,[\,\mathbf{E}[\,Y|X\,]\,] = \mathbf{E}\Big[\,\mathbf{E}[\,Y|X\,]^2\,\Big] - \mathbf{E}[\,Y\,]^2 \tag{3}$$

By adding equations (2) and (3) we get the result.

For the second part, let X be the number of balls that are assigned to the first bin. We compute $\mathbf{Var}\,[\,X\,] = \mathbf{E}[\,\mathbf{Var}\,[\,X|M\,]\,] + \mathbf{Var}\,[\,\mathbf{E}[\,X|M\,]\,]$. In lecture 6 slide 24 we computed $\mathbf{E}[\,X|M\,] = \sum_{i=1}^{\infty} \mathbf{1}_{\{i\le M\}} = M/n$. Moreover, by definition of conditional variance, we get

$$\mathbf{Var}\,[\,X|M\,] = \mathbf{E}\Big[\,(X-M/n)^2\Big|M\,\Big] \overset{p3,p5}{=} \mathbf{E}\big[\,X^2|M\,\big] - 2(M/n)\mathbf{E}[\,X|M\,] + (M/n)^2$$
$$= \mathbf{E}\big[\,X^2|M\,\big] - (M/n)^2$$

We just need to compute $\mathbf{E}\left[X^2 | M \right]$. Let X_i be 1 if ball number i is assigned to the first bin, otherwise X_i is 0. Then $X = \sum_{i=1}^{\infty} X_i \mathbf{1}_{\{i \leq M\}}$ and therefore

$$X^2 = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} X_i X_j \mathbf{1}_{\{i,j \leq M\}}$$

Note that $X_i^2 = X_i$, then

$$\mathbf{E}\left[X^2 | M \right] = \sum_{i=1}^{\infty} \mathbf{E}[X_i | M] + 2 \sum_{1 = i < j < \infty} \mathbf{E}[X_i X_j | M] \, \mathbf{1}_{\{i,j \leq M\}}$$

Finally, use that the location of a ball is independent of how many balls we assigned in total. Therefore $\mathbf{E}[X_i | M] = 1/n$ and $\mathbf{E}[X_i X_j | M] = 1/n^2$ for $i \neq j$. We conclude that

$$\mathbf{Var}\left[X | M \right] = M/n + M(M-1)/n^2 - (M/n)^2 = M/n - M/n^2$$

and $\mathbf{E}[\mathbf{Var}\left[X | M \right]] = (\mu/n) \left(1 - \frac{1}{n}\right)$

On the other hand, remember that $\mathbf{E}[X | M] = M/n$. Then

$$\mathbf{Var}\left[\mathbf{E}[X | M] \right] = \mathbf{Var}\left[M/n \right] = \frac{1}{n^2} \mathbf{Var}\left[M \right] = \sigma^2/n^2.$$

By adding $\mathbf{Var}\left[\mathbf{E}[X | M] \right]$ and $\mathbf{E}[\mathbf{Var}\left[X | M \right]]$ we get the result.

Question 6. *Consider a coin that shows head with probability p. What is the expected number of flips required to observe a run of n consecutive heads?*

Solution: Done in exercise class.

Question 7. *Let $X_1, \ldots, X_n$ i.i.d. samples from a distribution of interest. We know that $\mathbf{E}[X_i] = \mu$ and $\mathbf{Var}\left[X_i \right] = \sigma^2$ for all i, but we do not know the exact values of μ nor σ^2. We are given the mission to find an estimate $\hat{\mu}$ of the actual mean μ. We want the estimate $\hat{\mu}$ to satisfy the (δ, ε) condition: given ε, we want that $\hat{\mu} \in [\mu - \varepsilon\sigma, \mu + \varepsilon\sigma]$ with probability at least $1 - \delta$. How many data points X_i do we need to build an estimator satisfying the (δ, ε) condition?*

- *In a first attempt we can just deliver $\hat{\mu} = \frac{\sum_{i=1}^{n} X_i}{n}$, nevertheless, we cannot guarantee a good behaviour of such estimator, as we do not have enough information to compute a Chernoff Bound for it.*

1. *Prove that with $m = \lceil \frac{10}{\varepsilon^2} \rceil$ data points, we have that $\hat{\mu}_m = \left(\sum_{i=1}^{m} X_i\right)/m$ satisfies the $(1/10, \varepsilon)$ condition.*
2. *Write an algorithm that uses at most $O\left(\frac{\log(\delta^{-1})}{\varepsilon^2}\right)$ data points to build an estimate of μ satisfying the (δ, ε) condition.*

Hint.

Q6: *Recall how we deduce the expectation of a geometric in class.*

Q7: *For 2. consider batches of size $m = \lceil \frac{10}{\varepsilon^2} \rceil$. What can you say about more than half of them?*

Solution: Done in exercise class.
