Stochastic Optimisation

Solutions to Problem Sheet 1

---

1. The generating function of X_1 is given by

$$G_1(z) = E\Big[z^{X_1}\Big] = \sum_{n=0}^{\infty} P(X_1 = n) z^n.$$

But X_1 is Poisson with parameter λ_1, so $P(X_1 = n) = \lambda_1^n e^{-\lambda_1}/n!$. Substituting this above, we get

$$G_1(z) = \sum_{n=0}^{\infty} \frac{(\lambda_1 z)^n}{n!} e^{-\lambda_1} = e^{\lambda_1 z} e^{-\lambda_1} = e^{\lambda_1(z-1)}.$$

Similarly, X_2 has generating function $G_2(z) = e^{\lambda_2(z-1)}$.

Now, we have for the generating function of $X = X_1 + X_2$ that

$$G(z) = E\Big[z^X\Big] = E\Big[z^{X_1} z^{X_2}\Big] = E\Big[z^{X_1}\Big] E\Big[z^{X_2}\Big] = G_1(z) G_2(z),$$

since X_1 and X_2 are independent random variables (and hence so are z^{X_1} and z^{X_2}). Substituting for G_1 and G_2, we find that $G(z) = e^{(\lambda_1+\lambda_2)(z-1)}$, which we recognise as the generating function of a Poisson random variable with parameter $\lambda_1 + \lambda_2$. This completes the proof.

2. (a) The moment generating function of T is given by

$$M_T(\theta) = \mathbb{E}[e^{\theta T}] = \int_0^{\infty} \mu e^{-\mu x} e^{\theta x} dx = \begin{cases} \frac{\mu}{\mu - \theta}, & \text{if } \theta < \mu, \\ +\infty, & \text{otherwise.} \end{cases}$$

Hence,

$$\mathbb{E}[T] = M_T'(0) = \frac{\mu}{(\mu-\theta)^2}\Big|_{\theta=0} = \frac{1}{\mu}.$$

Next, recall that the cdf of an Exp(λ) random variable is $1 - e^{-\lambda x}$ for $x \geq 0$ and 0 for $x < 0$. Defining $Y = \mu T$, we see that

$$\mathbb{P}(Y > y) = \mathbb{P}(\mu T > y) = \mathbb{P}\big(T > \frac{y}{\mu}\big) = \exp\Big(-\mu \frac{y}{\mu}\Big) = e^{-y},$$

i.e., Y has the cdf of an Exp(1) random variable.

(b) By the conditional probability formula, we have for all $t, u \geq 0$ that

$$P(T > t+u | T > u) = \frac{P(\{T > t+u\} \cap \{T > u\})}{T > u} = \frac{P(T > t+u)}{T > u},$$

since the event $T > t+u$ is a subset of the event $T > u$, and hence their intersection is the event $T > t+u$. Now, recall that since T is exponentially distributed with parameter μ, $P(T > t) = e^{-\mu t}$ for all $t \geq 0$. Substituting this above,

$$P(T > t+u | T > u) = \frac{\exp(-\mu(t+u))}{\exp(-\mu u)} = e^{-\mu t} = P(T > t).$$

3. We have by Chernoff's bound that, for all $\theta \geq 0$,

$$\begin{aligned}
\mathbb{P}(X_1 + \ldots + X_n > nq) &\leq e^{-\theta nq} \mathbb{E}[\exp(\theta(X_1 + \ldots + X_n)] \\
&= e^{-\theta nq} \prod_{i=1}^{n} \mathbb{E}[e^{\theta X_i}] \quad \text{since the } X_i \text{ are mutually independent} \\
&= e^{-\theta nq} \big(1 - p + pe^{\theta}\big)^n.
\end{aligned}$$

Taking logarithms (which are a montone increasing function and preserve inequalities), we get

$$\log \mathbb{P}(X_1 + \ldots + X_n > nq) \leq -\theta nq + n \log\big(1 - p + pe^{\theta}\big), \quad \forall\, \theta \geq 0. \tag{1}$$

We seek the minimum of the RHS over $\theta \geq 0$. Setting the derivative with respect to θ to 0, we obtain the equation

$$-nq + \frac{npe^{\theta}}{1 - p + pe^{\theta}} = 0,$$

from which it follows that

$$\frac{q}{1-q} = \frac{pe^{\theta}}{1-p}, \text{ i.e., } e^{\theta} = \frac{q(1-p)}{p(1-q)}.$$

Since $q > p$ by assumption, $q(1-p) > p(1-q)$, and it follows that the solution for θ is positive. Substituting this value of θ in (1), we get

$$\begin{aligned}
\log \mathbb{P}(X_1 + \ldots + X_n > nq) &\leq -nq \log \frac{q(1-p)}{p(1-q)} + n \log\Big((1-p)\Big(1 + \frac{q}{1-q}\Big)\Big) \\
&= -nBbigl(q \log \frac{q}{p} - q \log \frac{1-q}{1-p} + \log \frac{1-p}{1-q}\Big) = -nK(q;p).
\end{aligned}$$

4. We have by Chernoff's bound that, for all $\theta \leq 0$,

$$\begin{aligned}
\mathbb{P}(X_1 + \ldots + X_n < n\mu) &\leq e^{-\theta n\mu} \mathbb{E}[\exp(\theta(X_1 + \ldots + X_n)] \\
&= e^{-\theta n\mu} \prod_{i=1}^{n} \mathbb{E}[e^{\theta X_i}] \quad \text{since the } X_i \text{ are mutually independent} \\
&= e^{-\theta n\mu} \big(\exp(\lambda(e^{\theta} - 1)\big)^n,
\end{aligned}$$

using the expression for the generating function of a Poisson random variable derived in Problem 1. Now, taking logarithms, we obtain

$$\frac{1}{n} \log \mathbb{P}(X_1 + \ldots + X_n < n\mu) \leq -\theta\mu + \lambda(e^{\theta} - 1) \quad \forall\, \theta \leq 0.$$

To minimise the expression on the RHS, we set its derivative with respect to θ to 0. This yields the equation $\mu = \lambda e^{\theta}$, and so $\theta = \log(\mu/\lambda)$. As μ was assumed to be smaller than λ, $\theta < 0$ as required. Substituting for θ above, we get

$$\frac{1}{n} \log \mathbb{P}(X_1 + \ldots + X_n < n\mu) \leq -\mu \log \frac{\mu}{\lambda} + \mu - \lambda = -I(\mu; \lambda),$$

as required.

5. Applying Chernoff's bound, we get that, for any $\theta \geq 0$,

$$\begin{aligned}\mathbb{P}(X_1 + \ldots + X_n > nx) &\leq e^{-\theta nx}\mathbb{E}[\exp(\theta(X_1 + \ldots + X_n)] \\ &= e^{-\theta nx}\prod_{i=1}^{n}\mathbb{E}[e^{\theta X_i}] \quad \text{since the } X_i \text{ are mutually independent} \\ &= e^{-\theta nx}\Big(\frac{\lambda}{\lambda-\theta}\Big)^n, \quad \theta \leq \lambda,\end{aligned}$$

using the expression for the mgf of an exponential random variable derived in Problem 2. Taking logarithms, we get

$$\frac{1}{n}\log \mathbb{P}(X_1 + \ldots + X_n > nx) \leq -\theta x + \log\frac{\lambda}{\lambda-\theta}, \quad \theta \in [0, \lambda).$$

We minimise the expression on the RHS by setting its derivative with respect to θ to 0. This yields $x = 1/(\lambda - \theta)$, i.e., $\theta = \lambda - 1/x$, which is positive if $x > 1/\lambda$, as we assumed. Substituting this in the bound, we get

$$\frac{1}{n}\log \mathbb{P}(X_1 + \ldots + X_n > nx) \leq -\lambda x + 1 + \log(\lambda x) = -J(x;\lambda).$$

6. (a) The mgf of Z is given by

$$\begin{aligned}M_Z(\theta) &= \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\Big(-\frac{x^2}{2} + \theta x\Big)dx \\ &= \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\Big(-\frac{(x-\theta)^2}{2} + \frac{\theta^2}{2}\Big)dx \\ &= e^{\theta^2/2}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\Big(-\frac{y^2}{2}\Big)dy = e^{\theta^2/2}.\end{aligned}$$

To obtain the third equality, we took the constant term $\exp(\theta^2/2)$ out of the integral, and made the change of variables $y = xi\theta$. To obtain the last equality, we recognised $(1/\sqrt{2\pi}\exp(-y^2/2)$ as the density of a standard normal random variable, which integrates to 1 as it is a probability density function.

(b) Notice that if $X_i \sim N(\mu, \sigma^2)$ and we define $Y_i = (X_i - \mu)/\sigma$, then $Y_i \sim N(0, 1)$. Moreover, the Y_i are also iid, and their mgf is given in part (a). Hence, we can use the Chernoff bound to write, for any $\theta \geq 0$, that

$$\begin{aligned}\mathbb{P}\Big(\sum_{i=1}^{n}X_i > n\gamma\Big) &\leq \mathbb{P}\Big(\sum_{i=1}^{n}Y_i > \frac{n(\gamma-\mu)}{\sigma}\Big) \\ &\leq \exp\Big(-\theta\frac{n(\gamma-\mu)}{\sigma}\Big)\mathbb{E}\Big[\exp\theta\sum_{i=1}^{n}Y_i\Big].\end{aligned}$$

Taking logarithms, using the independence of the Y_i and the expression for their mgf computed in part (a), we get

$$\frac{1}{n}\log \mathbb{P}\Big(\sum_{i=1}^{n}X_i > n\gamma\Big) \leq -\theta\frac{\gamma-\mu}{\sigma} + \frac{\theta^2}{2}, \quad \forall\,\theta \geq 0.$$

The mnimum of the expression on the RHS over $\theta \geq 0$ is attained at $\theta = (\gamma - \mu)/\sigma$, leading us to conclude that

$$\frac{1}{n} \log \mathbb{P}\Big(\sum_{i=1}^{n} X_i > n\gamma\Big) \leq -\frac{(\gamma - \mu)^2}{2\sigma^2},$$

as claimed in the question.

7. Define $Y = \langle \boldsymbol{\theta}, \mathbf{X} \rangle$. Using standard properties of inner products, $\eta Y = \langle \eta\boldsymbol{\theta}, \mathbf{X} \rangle$. Now,

$$\mathbf{X} \in H(\boldsymbol{\theta}, y) \;\Rightarrow\; \langle \boldsymbol{\theta}, \mathbf{X} \rangle \geq y \;\Rightarrow\; \langle \eta\boldsymbol{\theta}, \mathbf{X} \rangle \geq \eta y, \quad \forall\, \eta \geq 0.$$

Consequently,

$$\mathbb{P}(\mathbf{X} \in H(\boldsymbol{\theta}, y)) \leq \mathbb{P}(\langle \eta\boldsymbol{\theta}, \mathbf{X} \rangle \geq \eta y) \leq e^{-\eta y} \mathbb{E}[\exp(\langle \eta\boldsymbol{\theta}, \mathbf{X} \rangle)],$$

by Markov's inequality.

8. We shall use Hoeffding's inequality, which states that, if Y_i are iid, take values in $[0, 1]$, and have mean ν, then

$$\mathbb{P}\Big(\sum_{i=1}^{n} (Y_i - \nu) > nt\Big) \leq \exp(-2nt^2).$$

Define $Y_i = (X_i - a)/(b - a)$ and $\nu = (\mu - a)(b - a)$, where $\mu = \mathbb{E}[X_1]$. Then it is clear that Y_i are iid, take values in $[0, 1]$ and have mean ν. We can also rewrite the event $\sum_{i=1}^{n} (X_i - \mu) > nt$ as the event

$$\Big\{\sum_{i=1}^{n} \frac{(X_i - a) - (\mu - a)}{b - a} > n\frac{t}{b - a}\Big\} = \Big\{\sum_{i=1}^{n} (Y_i - \nu) > n\frac{t}{b - a}\Big\}.$$

The result now follows from Hoeffding's inequality.