Lecturer: Prof. Dr. Mete SONER
Coordinator: Yilin WANG

ETH Zürich

Solution Series 4

Q1. **(a)** Take $p \in [0,1]$ and $n \in \mathbb{N}\backslash\{0\}$. We say that $X \sim Bin(n,p)$ if the distribution of X is

$$\mathbb{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k \in \{0,1,...,n\}.$$

Show that this is indeed a probability distribution using 2 different methods:

i. Calculating $\sum_k \mathbb{P}(X=k)$.

ii. Representing this probability in terms of the box model with replacement.

(b) Take $K, n \in \mathbb{N}$ and $N \in \mathbb{N}\backslash\{0\}$ with $K, n \leq N$. We say that a random variable $X \sim Hyp(N,K,n)$ if its distribution is given by

$$\mathbb{P}(X=k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} \qquad k \in \{\max\{0, n+K-N\}, .., \min\{n,K\}\}$$

Show that this is indeed a probability distribution using 2 different methods:

i. Calculating $\sum_k \mathbb{P}(X=k)$.
Hint: Calculate $(1+x)^n$ in two different ways and identify the terms.

ii. Representing this probability in the box model without replacement.

Solution

(a) Probability:

i.

$$\begin{aligned}\sum_{k=0}^{n} \mathbb{P}(X=k) &= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \\ &= (p + (1-p))^n = 1.\end{aligned}$$

ii. If we have an urn with replacement with r red balls and b blue and we draw a ball n times, we have that

$$\begin{aligned}\mathbb{P}(\{\text{There are } k \text{ red } n-k \text{ blue }\}) &= \frac{|\{\text{There are } k \text{ red and } n-k \text{ blue}\}|}{|\{\text{Possible results}\}|} \\ &= \frac{\binom{n}{k} r^k b^{n-k}}{(r+b)^n} \\ &= \binom{n}{k}\left(\frac{r}{r+b}\right)^k \left(\frac{b}{r+b}\right)^{n-k} \\ &= \binom{n}{k} p^k (1-p)^{n-k},\end{aligned}$$

with $p = \frac{r}{r+b}$. Given that in every experiment we draw $k \in \{0, .., n\}$ red balls makes the expression a probability measure, i.e.,

$$\begin{aligned} 1 &= \mathbb{P}\left(\bigcup_{k=0}^{n} \{\text{We draw } k \text{ red bals}\}\right) \\ &= \sum_{k=1}^{n} \mathbb{P}\left(\{\text{We draw } k \text{ red bals}\}\right) \\ &= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k}. \end{aligned}$$

(b) Probability:

i. Note that

$$\begin{aligned} \sum_{n=0}^{N} \binom{N}{n} x^n &= (1+x)^N \\ &= (1+x)^K (1+x)^{N-K} \\ &= \sum_{k=0}^{K} \binom{K}{k} x^k \sum_{j=0}^{N-k} \binom{N-K}{j} x^j \\ &= \sum_{k=0}^{K} \sum_{j=0}^{N-K} \binom{K}{k} \binom{N-K}{j} x^{k+j} \\ &= \sum_{n=0}^{N} \sum_{k=\max\{0,K+n-N\}}^{\min\{n,K\}} \binom{K}{k} \binom{N-K}{n-k} x^n, \end{aligned}$$

where we made the change of variables $n = k + j$. Given that two polynomials are equal iff all of its coefficients are equal we have that

$$\begin{aligned} &\sum_{k=\max\{0,K+n-N\}}^{\min\{n,K\}} \binom{K}{k} \binom{N-K}{n-k} = \binom{N}{n} \\ \Rightarrow &\sum_{k=\max\{0,K+n-N\}}^{\min\{n,K\}} \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} = 1, \end{aligned}$$

so it is a probability measure.

ii. If you have N balls K of which are red and $N - K$ blue and you are drawing them out without replacement. We have that the event $B :=$ "in the n-th draw we have extracted k balls red and $n - k$ balls blue" is given by

$$\begin{aligned} P(B) &= \frac{|\{\text{Ways of taking out } k \text{ balls red and } n-k \text{ blue }\}|}{|\{\text{Ways of taking out n balls}\}|} \\ &= \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}. \end{aligned}$$

Given that in every experiment we extract $k \in \{0, .., n\}$ red balls, the expression is a probability measure.

Q2. Let X_1 and X_2 follow a normal distribution with mean μ_i and variance σ_i^2. Prove that if X_1 is independent of X_2 then $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Solution: Let's suppose, first, that $\mu_1 = \mu_2 = 0$. We just have to use the convolution formula:

$$\begin{aligned}
f_{X_1+X_2}(x) &= \frac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{\infty}\exp(-\frac{y^2}{2\sigma_1^2})\exp(-\frac{(x-y)^2}{2\sigma_2^2})dy \\
&= \frac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{\infty}\exp\left(-\frac{\sigma_2^2y^2+\sigma_1^2(x-y)^2}{2\sigma_1^2\sigma_2^2}\right)dy \\
&= \frac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{\infty}\exp\left(-\frac{\sigma_2^2y^2+\sigma_1^2x^2+\sigma_1^2y^2-2\sigma_1^2xy}{2\sigma_1^2\sigma_2^2}\right)dy \\
&= \frac{\exp(-\frac{x^2}{2\sigma_2^2})}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{\infty}\exp\left(-(\sigma_1^2+\sigma_2^2)\frac{y^2-\frac{2\sigma_1^2xy}{(\sigma_1^2+\sigma_2^2)}}{2\sigma_1^2\sigma_2^2}\right)dy \\
&= \frac{\exp(-\frac{x^2}{2\sigma_2^2}+\frac{\sigma_1^2x^2}{2\sigma_2^2(\sigma_1^2+\sigma_2^2)})}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{\infty}\exp\left(-\frac{\left(y-\frac{\sigma_1^2x}{(\sigma_1^2+\sigma_2^2)}\right)^2}{\frac{2\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}}\right)dy \\
&= \frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2+\sigma_2^2}}\exp(-\frac{x^2}{2(\sigma_1^2+\sigma_2^2)}),
\end{aligned}$$

that is the distribution function of a normal random variable with parameter $N(0, \sigma_1^2 + \sigma_2^2)$.

For the general case note, that $X_i - \mu_i$ is distributed as $N(0, \sigma_i^2)$. So $(X_1 - \mu_1) + (X_2 - \mu_2) \sim N(0, \sigma_1^2 + \sigma_2^2)$, then $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Q3. Let X be a standard normal random variable.

(a) Prove that if we take $Y := X^2$, then $f_Y(y) = ce^{-y/2}y^{-1/2}\mathbf{1}_{\{y\geq 0\}}$ (We say that Y is distributed according to a χ-squared with one degree of freedom).

(b) If Y_1 and Y_2 are two independent copies of Y, prove that $f_{Y_1+Y_2}(x) = c_2e^{-x/2}\mathbf{1}_{\{x\geq 0\}}$. What is the name of this distribution.

(c) With the help of induction prove that $\sum_{i=1}^n Y_i$, where $(Y_i)_{i=1}^n$ are independent copies of Y, has as a density function

$$f_{\sum_{i=1}^n Y_i}(x) = c_n x^{\frac{n}{2}-1}e^{-\frac{x}{2}}\mathbf{1}_{\{x\geq 0\}}.$$

This is call a χ-squared distribution with n degrees of freedom.

Solution:

(a) We have that the cumulative distribution function of Y for $y \geq 0$ is given by:

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y}) = 2F_X(\sqrt{y}) - 1.$$

Then, taking the derivative we have:

$$f_Y(y) = f_x(\sqrt{y})y^{-\frac{1}{2}}\mathbf{1}_{\{y\geq 0\}} = ce^{-\frac{y}{2}}y^{-\frac{1}{2}}\mathbf{1}_{\{y\geq 0\}}.$$

(b) By the convolution formula we have that:

$$\begin{aligned}
f_{Y_1+Y_2}(x) &= \int_0^x f_Y(x-y)f_Y(y)dy\mathbf{1}_{\{x\geq 0\}} \\
&= c_1^2\int_0^x (x-y)^{-\frac{1}{2}}e^{-\frac{x-y}{2}}y^{-\frac{1}{2}}e^{-\frac{y}{2}}dy\mathbf{1}_{\{x\geq 0\}} \\
&= c_1^2 e^{-\frac{x}{2}}\int_0^x (x-y)^{-\frac{1}{2}}y^{-\frac{1}{2}}dy\mathbf{1}_{\{x\geq 0\}} \\
&= c_1^2\left(\int_0^1 x(x-ux)^{-\frac{1}{2}}(ux)^{-\frac{1}{2}}du\right)e^{-\frac{x}{2}}\mathbf{1}_{\{x\geq 0\}} \\
&= \left(c_1^2\int_0^1 (1-u)^{-\frac{1}{2}}u^{-\frac{1}{2}}du\right)e^{-\frac{x}{2}}\mathbf{1}_{\{x\geq 0\}}.
\end{aligned}$$

This distribution is that of an exponential random variable.

(c) It's clear that the base case is true, now let's prove the inductive step. Suppose that the proposition is true for $n-1$, then

$$\begin{aligned}
f_{\sum_{i=1}^n Y_i}(x) &= \int_0^x f_Y(x-y)f_{\sum_{i=1}^{n-1} Y_i}(y)dy\mathbf{1}_{\{x\geq 0\}} \\
&= c_1c_{n-1}\int_0^x (x-y)^{-\frac{1}{2}}e^{-\frac{x-y}{2}}y^{\frac{n-1}{2}-1}e^{-\frac{y}{2}}dy\mathbf{1}_{\{x\geq 0\}} \\
&= c_1c_{n-1}e^{-\frac{x}{2}}\int_0^1 (x-xu)^{-\frac{1}{2}}(xu)^{\frac{n-1}{2}-1}xdu\mathbf{1}_{\{x\geq 0\}} \\
&= \left(c_1c_{n-1}\int_0^1 (1-u)^{-\frac{1}{2}}(u)^{-\frac{n-1}{2}-1}du\right)e^{-\frac{x}{2}}x^{\frac{n}{2}-1}\mathbf{1}_{\{x\geq 0\}}.
\end{aligned}$$

Q4. Memorylessness of exponential random variable. We say that a random variable X has an exponential distribution of parameter λ (write it $\mathcal{E}(\lambda)$) if for all $t \geq 0$:

$$\mathbb{P}(X \geq t) = e^{-\lambda t}.$$

(a) Find the density function (with respect to the Lebesgue Measure) of an exponential random variable.

(b) Show that if $X_1 \sim \mathcal{E}(\lambda_1)$, $X_2 \sim \mathcal{E}(\lambda_2)$ and $X_1 \perp X_2$, then $\min\{X_1, X_2\} \sim \mathcal{E}(\lambda_1 + \lambda_2)$.

(c) Show that

$$\mathbb{P}\left(X \geq t+h \mid X \geq h\right) = \mathbb{P}(X \geq t).$$

This property is called memorylessness. We want to prove that the only random variable that has the memorylessness property is the exponential random variable. Suppose that $Y : \Omega \mapsto \mathbb{R}^+$ has the memorylessness property, i.e.,

$$\mathbb{P}\left(Y \geq t+h \mid Y \geq h\right) = \mathbb{P}(X \geq t).$$

(d) Define $G(t) := \mathbb{P}(Y \geq t)$ and prove that $G(t+h) = G(t)G(h)$.

(e) Prove that for all $m, n \in \mathbb{N}$, $G\left(\frac{m}{n}\right) = G(1)^{\frac{m}{n}}$.

(f) Using the monotone property of G prove that for all $t \geq 0$, $G(t) = G(1)^t$. Conclude that Y has an exponential distribution and make explicit the parameter.

Solution

(a) To find the density we just have to derive the CDF

$$F(t) := \mathbb{P}(X \leq t) = 1 - \mathbb{P}(X \geq t) = 1 - e^{-\lambda t}.$$

Then its density is

$$f(t) := F'(t) = \lambda e^{-\lambda t}.$$

(b) We just have to compute

$$\begin{aligned}
\mathbb{P}\left(\min\{X_1, X_2\} > t\right) &= \mathbb{P}(X_1 > t, X_2 > t) \\
&= \mathbb{P}(X_1 > t)\mathbb{P}(X_2 > t) \\
&= e^{-\lambda_1 t} e^{-\lambda_2 t} \\
&= e^{-(\lambda_1 + \lambda_2)t}.
\end{aligned}$$

This is the definition of $\min\{X_1, X_2\} \sim \mathcal{E}(\lambda_1 + \lambda_2)$.

(c) We just have to compute

$$\mathbb{P}\left(Y \geq t+h \mid Y \geq h\right) = \frac{\mathbb{P}(Y \geq t+h)}{\mathbb{P}(Y \geq h)} = e^{-\lambda t} = \mathbb{P}(Y \geq t).$$

(d) We have to compute

$$\begin{aligned}
G(t+h) &= \mathbb{P}(Y \geq t+h) \\
&= \frac{\mathbb{P}(Y \geq t+h)}{\mathbb{P}(Y \geq h)} \mathbb{P}(Y \geq h) \\
&= \mathbb{P}(Y \geq t+h \mid Y \geq h)\mathbb{P}(Y \geq h) \\
&= G(t)G(h).
\end{aligned}$$

(e) First we will prove by induction that for all $n \in \mathbb{N}$ and $(a_n)_{n\in\mathbb{N}} \subseteq \mathbb{R}$ we have that $G(\sum_{i=1}^n a_i) = \prod_{i=1}^n G(a_i)$. It's clear when $n = 1$, then assuming it's true for n

$$G\left(\sum_{i=1}^{n+1} a_i\right) = G(a_{n+1})G\left(\sum_{i=1}^{n} a_n\right) = \prod_{i=1}^{n+1} G(a_i).$$

Take $m, n \in \mathbb{N}$, we have that

$$\begin{aligned} G(1)^m &= G\left(\sum_{i=1}^{m} 1\right) = G\left(\sum_{i=1}^{n} \frac{m}{n}\right) = G\left(\frac{m}{n}\right)^n \\ \Rightarrow G(1)^{\frac{m}{n}} &= G\left(\frac{m}{n}\right) \end{aligned}$$

(f) Finally, take $t \in \mathbb{R}^+$ and $(t_n)_{n\in\mathbb{N}}, (s_n)_{n\in\mathbb{N}} \subseteq \mathbf{Q}$ so that $t_n \nearrow t$ and $s_n \searrow t$. Then thanks to the monotonicity of $G(t)$

$$\begin{aligned} &G(t_n) \leq G(t) \leq G(s_n) \\ \Rightarrow &G(1)^{t_n} \leq G(t) \leq G(1)^{s_n} \\ \Rightarrow &G(t) = G(1)^t. \end{aligned}$$

Finally we have that $\mathbb{P}(Y \geq t) = G(1)^t = e^{-\ln\left(\frac{1}{G(1)}\right)t}$, then $Y \sim \mathcal{E}\left(\ln\left(\frac{1}{G(1)}\right)\right)$.