ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Chair for Mathematical Information Science

Prof. Dr. H. Bölcskei
Sternwartstrasse 7
CH-8092 Zürich

Mathematics of Information

Spring semester 2022

Solutions to Problem Set 11

Problem 1 Integrated formula for the expectation.

We have the following chain of equalities

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}\left[\int_0^X 1\,ds\right]\\ &= \mathbb{E}\left[\int_0^\infty \mathbb{1}_{0\le s\le X}\,ds\right]\\ &= \int_0^\infty \mathbb{E}\left[\mathbb{1}_{X\ge s}\right]\,ds\\ &= \int_0^\infty \mathbb{P}\left[X\ge s\right]\,ds,\end{aligned}$$

where the third equality is a direct application of Fubini's theorem.

Problem 2 Median vs Mean.

a) Take any $c\in\mathbb{R}$, then

$$\begin{aligned}\mathbb{E}\left[|X-c|^2\right] &= \mathbb{E}\left[|X-\mathbb{E}[X]+\mathbb{E}[X]-c|^2\right]\\ &= \mathbb{E}\left[|X-\mathbb{E}[X]|^2\right]+\mathbb{E}\left[|c-\mathbb{E}[X]|^2\right]+2\mathbb{E}\left[(X-\mathbb{E}[X])(\mathbb{E}[X]-c)\right]\\ &= \mathbb{E}\left[|X-\mathbb{E}[X]|^2\right]+|c-\mathbb{E}[X]|^2+2\mathbb{E}\left[X-\mathbb{E}[X]\right](\mathbb{E}[X]-c)\\ &= \mathbb{E}\left[|X-\mathbb{E}[X]|^2\right]+|c-\mathbb{E}[X]|^2\ge\mathbb{E}\left[|X-\mathbb{E}[X]|^2\right],\end{aligned}$$

with equality if and onli if $c=\mathbb{E}[X]$. This is the desired result.

b) Consider the function f defined on $\mathbb{R}$ by

$$f(c) := \mathbb{E}\left[|X-c|\right].$$

Using the integrated formula for the expectation derived in Problem 1, we obtain

$$\begin{aligned}f(c) &= \int_0^\infty \mathbb{P}\left[|X-c|\ge t\right]dt\\ &= \int_0^\infty \mathbb{P}\left[X\ge t+c\right]+\mathbb{P}\left[X\le c-t\right]dt\\ &= \int_c^\infty \mathbb{P}\left[X\ge s\right]ds+\int_{-\infty}^c \mathbb{P}\left[X\le s\right]ds,\end{aligned}$$

for all $c \in \mathbb{R}$. This implies

$$\begin{aligned} f(c) &= \int_c^{\infty} \mathbb{P}\left[X \geq s\right] ds + \int_{-\infty}^{c} \mathbb{P}\left[X \leq s\right] ds \\ &= \int_M^{\infty} \mathbb{P}\left[X \geq s\right] ds + \int_c^{M} \mathbb{P}\left[X \geq s\right] ds + \int_{-\infty}^{M} \mathbb{P}\left[X \leq s\right] ds + \int_M^{c} \mathbb{P}\left[X \leq s\right] ds \\ &= f(M) + \int_c^{M} \mathbb{P}\left[X \geq s\right] - \mathbb{P}\left[X \leq s\right] ds \geq f(M), \end{aligned}$$

where the last inequality uses that if $c \leq s \leq M$ the integrand is nonnegative and if $M \leq s \leq c$ the integrand is nonpositive. In particular, we have proven that M is a minimum of f, which concludes the proof.

c) We have the following inequalities

$$\begin{aligned} |\mathbb{E}\left[X\right] - M| &= |\mathbb{E}\left[X - M\right]| \\ &\leq \mathbb{E}\left[|X - M|\right] \\ &\leq \mathbb{E}\left[|X - \mathbb{E}[X]|\right] \\ &\leq \mathbb{E}\left[|X - \mathbb{E}[X]|^2\right]^{1/2} \\ &= \sqrt{\mathbb{V}[X]}, \end{aligned}$$

where the first and the last inequalities come from Jensen's inequality, and the second inequality is obtained by using that M minimizes the L^1 loss. This is the desired result.

Problem 3 Properties of sub-Gaussian random variables.

a) If we perform a Taylor expansion of

$$\mathbb{E}\left[e^{\lambda(X - \mathbb{E}[X])}\right] \leq e^{\frac{\sigma^2\lambda^2}{2}}$$

to the second order in $\lambda > 0$, we obtain

$$\mathbb{E}\left[1 + \lambda(X - \mathbb{E}[X]) + \frac{\lambda^2}{2}(X - \mathbb{E}[X])^2 + O(\lambda^3)\right] \leq 1 + \frac{\sigma^2\lambda^2}{2} + O(\lambda^3),$$

which, after expanding the LHS, yields

$$1 + \lambda \underbrace{\mathbb{E}[X - \mathbb{E}[X]]}_{=0} + \frac{\lambda^2}{2}\mathbb{V}[X] + O(\lambda^3) \leq 1 + \frac{\sigma^2\lambda^2}{2} + O(\lambda^3).$$

Simplifying and taking the limit $\lambda \to 0$ gives the desired result.

b) The result is obvious if $s = 0$, so we may take $0 < s < 1$. We follow the hint and multiply both sides by $e^{-\frac{\lambda^2\sigma^2}{2s}}$ to obtain

$$\mathbb{E}\left[e^{\lambda X - \frac{\lambda^2\sigma^2}{2s}}\right] \leq e^{\frac{\lambda^2\sigma^2(s-1)}{2s}}.$$

If we integrate the RHS with respect to λ, we get

$$\int_{\mathbb{R}} e^{\frac{\lambda^2\sigma^2(s-1)}{2s}} d\lambda = \sqrt{\frac{2\pi s}{\sigma^2(1-s)}} \int_{\mathbb{R}} \left(\sqrt{\frac{2\pi s}{\sigma^2(1-s)}}\right)^{-1} e^{-\frac{\lambda^2\sigma^2(1-s)}{2s}} d\lambda$$
$$= \sqrt{\frac{2\pi s}{\sigma^2(1-s)}},$$

where in the last equality we used that the integrand is the density of a Gaussian distribution and therefore integrates to one. If we integrate the LHS and we apply Fubini to swap the integral and the expectation, we obtain

$$\int_{\mathbb{R}} \mathbb{E}\left[e^{\lambda X - \frac{\lambda^2\sigma^2}{2s}}\right] d\lambda = \mathbb{E}\left[\int_{\mathbb{R}} e^{\lambda X - \frac{\lambda^2\sigma^2}{2s}} d\lambda\right].$$

We now observe that, for all $x \in \mathbb{R}$,

$$\int_{\mathbb{R}} e^{\lambda x - \frac{\lambda^2\sigma^2}{2s}} d\lambda = \int_{\mathbb{R}} e^{-\left(\frac{\sigma}{\sqrt{2s}}\lambda - \sqrt{\frac{s}{2\sigma^2}}x\right)^2 + \frac{s}{2\sigma^2}x^2} d\lambda$$
$$= e^{\frac{s}{2\sigma^2}x^2} \sqrt{\frac{2\pi s}{\sigma^2}} \int_{\mathbb{R}} \left(\sqrt{\frac{2\pi s}{\sigma^2}}\right)^{-1} e^{-\left(\frac{\sigma}{\sqrt{2s}}\lambda - \sqrt{\frac{s}{2\sigma^2}}x\right)^2} d\lambda$$
$$= e^{\frac{s}{2\sigma^2}x^2} \sqrt{\frac{2\pi s}{\sigma^2}},$$

where we again used that the density of a Gaussian random variable integrates to one. Combining both of the results, we get

$$\mathbb{E}\left[e^{\frac{s}{2\sigma^2}X^2} \sqrt{\frac{2\pi s}{\sigma^2}}\right] \leq \sqrt{\frac{2\pi s}{\sigma^2(1-s)}},$$

which, after simplification, yields the desired result

$$\mathbb{E}\left[e^{\frac{sX^2}{2\sigma^2}}\right] \leq \frac{1}{\sqrt{1-s}}.$$

Problem 4 Bounded random variables are sub-Gaussian.

a) By evaluating ψ and ψ' at 0, we have

$$\psi(0) = \log \mathbb{E}[1] = 0,$$

and

$$\psi'(0) = \frac{\mathbb{E}\left[X e^{0 \cdot X}\right]}{\mathbb{E}\left[e^{0 \cdot X}\right]} = \mathbb{E}[X] =: \mu.$$

b) We compute

$$\psi''(\lambda) = \left(\frac{\mathbb{E}\left[Xe^{\lambda X}\right]}{\mathbb{E}\left[e^{\lambda X}\right]}\right)'$$
$$= \frac{\mathbb{E}\left[X^2 e^{\lambda X}\right]}{\mathbb{E}\left[e^{\lambda X}\right]} - \frac{\mathbb{E}\left[Xe^{\lambda X}\right]^2}{\mathbb{E}\left[e^{\lambda X}\right]^2}.$$

Let $\mathbb{P}$ denote the probability distribution of X. We then observe that ψ'' takes the form of a variance with respect to a 'tilted' distribution, namely $\psi''(\lambda) = \mathrm{Var}_\lambda(X)$ where the variance Var_λ is taken relatively to the distribution $\mathbb{Q}_\lambda$ of X defined by

$$d\mathbb{Q}_\lambda(x) = \frac{e^{\lambda x}}{\mathbb{E}_{\mathbb{P}}\left[e^{\lambda X}\right]} d\mathbb{P}(x).$$

For any real-valued random variable Y, such that $Y \in [a, b]$ almost surely, its variance is bounded as

$$\begin{aligned}
\mathrm{Var}[Y] &= \inf_t \mathbb{E}\left[(Y - t)^2\right] \\
&\leq \mathbb{E}\left[\left(Y - \frac{a+b}{2}\right)^2\right] \\
&\leq \max\left\{\left(a - \frac{a+b}{2}\right)^2, \left(b - \frac{a+b}{2}\right)^2\right\} \\
&= \frac{(b-a)^2}{4}.
\end{aligned}$$

Applying this upper-bound to X yields the desired result:

$$\sup_{\lambda \in \mathbb{R}} |\psi''(\lambda)| \leq \frac{(b-a)^2}{4}.$$

c) Fix $\lambda \in \mathbb{R}$, by Taylor expansion, we have $\tilde{\lambda} \in \mathbb{R}$ with $|\tilde{\lambda}| \leq |\lambda|$ such that

$$\begin{aligned}
\log \mathbb{E}\left[e^{\lambda(X-\mu)}\right] &= \psi(\lambda) - \lambda\mu \\
&= \psi(\lambda) - \psi(0) - \lambda\psi'(0) \\
&= \frac{\lambda^2}{2}\psi''\left(\tilde{\lambda}\right) \\
&\leq \frac{\lambda^2[(b-a)^2/4]}{2}.
\end{aligned}$$

Exponentiating on both sides yields the desired result.

Problem 5 Convergence of random variables.

a) i) By definition of Ω_n, we have

$$\begin{aligned}\Omega_n &= \bigcup_{m\geq n} \{\omega \,|\, |X_m(\omega) - X(\omega)| > \varepsilon\} \\ &\subseteq \bigcup_{m\geq n} \{\omega \,|\, |X_m(\omega) - X(\omega)| > \varepsilon\} \cup \{\omega \,|\, |X_{n-1}(\omega) - X(\omega)| > \varepsilon\} \\ &= \bigcup_{m\geq n-1} \{\omega \,|\, |X_m(\omega) - X(\omega)| > \varepsilon\} \\ &= \Omega_{n-1}.\end{aligned}$$

Therefore $\{\Omega_n\}_{n\in\mathbb{N}}$ is decreasing and the limit set contains all the elements that are in all the sets Ω_n:

$$\Omega_\infty := \bigcap_{n\geq 1} \Omega_n.$$

ii) By definition of Ω_n, we have

$$\begin{aligned}\mathbb{P}\left[|X_n - X| > \varepsilon\right] &= \mathbb{P}\left[\{\omega \,|\, |X_n(\omega) - X(\omega)| > \varepsilon\}\right] \\ &\leq \mathbb{P}\left[\Omega_n\right].\end{aligned}$$

Since $\{\Omega_n\}_{n\in\mathbb{N}}$ is a monotone sequence of events with limit Ω_∞, the sequence $\{\mathbb{P}\left[\Omega_n\right]\}_{n\in\mathbb{N}}$ is also monotone with limit $\mathbb{P}\left[\Omega_\infty\right]$. Therefore, taking the limits yields the desired result:

$$\lim_{n\to\infty} \mathbb{P}\left[|X_n - X| > \varepsilon\right] \leq \mathbb{P}\left[\Omega_\infty\right].$$

iii) By definition of Ω_∞, we have

$$\begin{aligned}\mathbb{P}\left[\Omega_\infty\right] &= \mathbb{P}\left[\bigcap_{n\geq 1}\bigcup_{m\geq n} \{\omega \,|\, |X_m(\omega) - X(\omega)| > \varepsilon\}\right] \\ &= \mathbb{P}\left[\bigcap_{n\geq 1}\left\{\omega \,|\, \sup_{m\geq n} |X_m(\omega) - X(\omega)| > \varepsilon\right\}\right] \\ &\leq \mathbb{P}\left[\left\{\omega \,|\, \inf_{n\geq 1}\sup_{m\geq n} |X_m(\omega) - X(\omega)| \geq \varepsilon\right\}\right] \\ &\overset{(i)}{=} \mathbb{P}\left[\limsup_{n\to\infty} |X_n - X| \geq \varepsilon\right] \\ &= 1 - \mathbb{P}\left[\limsup_{n\to\infty} |X_n - X| < \varepsilon\right] \\ &\overset{(ii)}{=} 0,\end{aligned}$$

where (i) holds by definition of the lim sup and (ii) comes from the almost sure convergence of X_n.

b) We show that X_n converges in probability to a random variable X but not almost surely.

First note that

$$\mathbb{P}[|X_n| > \varepsilon] = \frac{1}{2^{\lfloor \log_2(n) \rfloor}}, \quad \forall \varepsilon \in (0, 1)$$

and

$$\mathbb{P}[|X_n| > \varepsilon] = 0, \quad \forall \varepsilon \geq 1.$$

Therefore, taking $X = 0$ almost surely, we have

$$\mathbb{P}[|X_n - X| > \varepsilon] \leq \frac{1}{2^{\lfloor \log_2(n) \rfloor}}, \quad \forall \varepsilon > 0,$$

from which we deduce that X_n converges in probability to X.

However, for all $\omega \in [0, 1]$ we have $\limsup_{n\to\infty} X_n(\omega) = 1$

$$\mathbb{P}\left[\limsup_{n\to\infty} |X_n - X| = 0\right] = 0 \neq 1.$$

X_n does not converge almost surely to X.