Foundations of Data Science
Gastpar, Urbanke

Ecole Polytechnique Fédérale, Lausanne: Fall 2024
September 10, 2024

Problem Set 1

For the Exercise Session on September 10

Last name	First name	SCIPER Nr	Points

Problem 1: Review of Random Variables

Let X and Y be discrete random variables defined on some probability space with a joint pmf $p_{XY}(x, y)$. Let $a, b \in \mathbb{R}$ be fixed.

(a) Prove that $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$. Do not assume independence.

(b) Prove that if X and Y are independent random variables, then $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$.

(c) Assume that X and Y are not independent. Find an example where $\mathbb{E}[X \cdot Y] \neq \mathbb{E}[X] \cdot \mathbb{E}[Y]$, and another example where $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$.

(d) Prove that if X and Y are independent, then they are also uncorrelated, i.e.,

$$Cov(X, Y) := \mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right] = 0. \tag{1}$$

(e) Find an example where X and Y are uncorrelated but dependent.

(f) Assume that X and Y are uncorrelated and let σ_X^2 and σ_Y^2 be the variances of X and Y, respectively. Find the variance of $aX + bY$ and express it in terms of $\sigma_X^2, \sigma_Y^2, a, b$.
Hint: First show that $Cov(X, Y) = \mathbb{E}[X \cdot Y] - \mathbb{E}[X] \cdot \mathbb{E}[Y]$.

Problem 2: Review of Gaussian Random Variables

A random variable X with probability density function

$$p_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}} \tag{2}$$

is called a *Gaussian* random variable.

(a) Explicitly calculate the mean $\mathbb{E}[X]$, the second moment $\mathbb{E}[X^2]$, and the variance $Var[X]$ of the random variable X.

(b) Let us now consider events of the following kind:

$$\Pr(X < \alpha). \tag{3}$$

Unfortunately for Gaussian random variables this cannot be calculated in closed form. Instead, we will rewrite it in terms of the standard Q-function:

$$Q(x) = \int_x^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}\, du \tag{4}$$

Express $\Pr(X < \alpha)$ in terms of the Q-function and the parameters m and σ^2 of the Gaussian pdf.

Like we said, the Q-function cannot be calculated in closed form. Therefore, it is important to have *bounds* on the Q-function. In the next 3 subproblems, you derive the most important of these bounds, learning some very general and powerful tools along the way:

(c) Derive the Markov inequality, which says that for any non-negative random variable X and positive a, we have

$$\Pr(X \geq a) \quad \leq \quad \frac{\mathbb{E}[X]}{a}. \tag{5}$$

(d) Use the Markov inequality to derive the Chernoff bound: the probability that a real random variable Z exceeds b is given by

$$\Pr(Z \geq b) \leq \mathbb{E}\big[e^{s(Z-b)}\big], \qquad s \geq 0. \tag{6}$$

(e) Use the Chernoff bound to show that

$$Q(x) \leq e^{-\frac{x^2}{2}} \quad \text{for } x \geq 0. \tag{7}$$

Problem 3: Moment Generating Function

In the class we had considered the logarithmic moment generating function

$$\phi(s) := \ln \mathbb{E}[\exp(sX)] = \ln \sum_x p(x) \exp(sx)$$

of a real-valued random variable X taking values on a finite set, and showed that $\phi'(s) = \mathbb{E}[X_s]$ where X_s is a random variable taking the same values as X but with probabilities $p_s(x) := p(x)\exp(sx)\exp(-\phi(s))$.

(a) Show that
$$\phi''(s) = \mathrm{Var}(X_s) := \mathbb{E}[X_s^2] - \mathbb{E}[X_s]^2$$
and conclude that $\phi''(s) \geq 0$ and the inequality is strict except when X is deterministic.

(b) Let $x_{\min} := \min\{x : p(x) > 0\}$ and $x_{\max} := \max\{x : p(x) > 0\}$ be the smallest and largest values X takes. Show that
$$\lim_{s \to -\infty} \phi'(s) = x_{\min}, \quad \text{and} \quad \lim_{s \to \infty} \phi'(s) = x_{\max}.$$

Problem 4: Hoeffding's Lemma

Prove Lemma 2.4 in the lecture notes. In other words, prove that if X is a zero-mean random variable taking values in $[a, b]$ then

$$\mathbb{E}[e^{\lambda X}] \leq e^{\frac{\lambda^2}{2}[(a-b)^2/4]}.$$

Expressed differently, X is $[(a-b)^2/4]$-subgaussian.

Hint: You can use the following steps to prove the lemma:

1. Let $\lambda > 0$. Let X be a random variable such that $a \leq X \leq b$ and $\mathbb{E}[X] = 0$. By considering the convex function $x \to e^{\lambda x}$, show that

$$\mathbb{E}[e^{\lambda X}] \leq \frac{b}{b-a} e^{\lambda a} - \frac{a}{b-a} e^{\lambda b}. \tag{8}$$

2. Let $p = -a/(b-a)$ and $h = \lambda(b-a)$. Verify that the right-hand side of (8) equals $e^{L(h)}$ where

$$L(h) = -hp + \log(1 - p + pe^h).$$

3. By Taylor's theorem, there exists $\xi \in (0, h)$ such that

$$L(h) = L(0) + hL'(0) + \frac{h^2}{2} L''(\xi).$$

   Show that $L(h) \leq h^2/8$ and hence $\mathbb{E}[e^{\lambda X}] \leq e^{\lambda^2 (b-a)^2/8}$.

Problem 5: Expected Maximum of Subgaussians

Let $\{X_i\}_{i=1}^n$ be a collection of n σ^2-subgaussian random variables, not necessarily independent of each other. Let $Y = \max_{i \in \{1,2,\cdots,n\}} X_i$. Prove that $\mathbb{E}[Y] \leq \sqrt{2\sigma^2 \log n}$. *Hint:* Recall that by Jensen, $e^{\lambda \mathbb{E}[X]} \leq \mathbb{E}[e^{\lambda X}]$.