# Stanford Stats 117 Final Examination

Closed book and one sheet of paper
Duration: 180 minutes

Fall 2024

**Name:**[1] ________________________________

**Student ID Number:** ________________________________

By taking this exam, you agree to be bound by the Stanford Honor Code, meaning specifically in this context that you

- will not give or receive aid in examinations;
- will do your share and take an active part in seeing to it that others as well as yourself uphold the spirit and letter of the Honor Code.

**Signature:** ________________________________

[1] It might be good to write your name on each page of your exam

**Question F.1:** We have a sensitive question we wish to ask an individual, whose answer $X$ is a binary random variable (where $X = 1$ corresponds to "Yes" and $X = 0$ corresponds to "No"). Instead of answering the question directly, the individual gives a *randomized response* $Z \in \{0, 1\}$, where for a value $0 \le p < \frac{1}{2}$, $Z$ has the distribution

$$Z = \begin{cases} 0 & \text{with probability } p \text{ independent of } X \\ 1 & \text{with probability } p \text{ independent of } X \\ X & \text{with probability } 1 - 2p. \end{cases}$$

(a) **(6 pts)** Give $\mathbb{E}[Z \mid X]$.

(b) **(6 pts)** Give scalars (real numbers) $a$ and $b$ such that $\mathbb{E}[aZ + b \mid X] = X$.

(c) **(6 pts)** Give $\mathrm{Var}(aZ + b)$ for your values of $a$ and $b$. (Our answer depends on $\mathrm{Var}(X)$.)

**Question F.2 (10 pts):** There are two assets, each of which has (random) returns $R_1$ and $R_2$, meaning that if you invest a dollar in asset $i$, at the end of the investment period it is worth $R_i$. The first asset guarantees constant return $R_1 = \mu_1$ with no variance. The other has expected return $\mu_2 = \mathbb{E}[R_2] > \mu_1$ and variance $\mathrm{Var}(R_2) = \sigma^2 > 0$. You invest a proportion $p$ of your wealth in asset 2, and the remaining proportion $1 - p$ in asset 1. (We allow $p > 1$ or $p < 0$, corresponding to borrowing against one of the assets.)

The total returns are thus $R = (1-p)R_1 + pR_2$. The *risk* of an investment is its variance; to choose an investment strategy (the proportion $p$), we maximize risk-adjusted returns

$$\mathbb{E}[R] - \lambda \cdot \mathrm{Var}(R), \tag{F.1}$$

where $\lambda > 0$ is a risk adjustment. For (arbitrary) values $\mu_1$, $\mu_2$, $\sigma^2 > 0$, and $\lambda > 0$, give the investment proportion $p$ that maximizes the risk-adjusted-return (F.1).

**Question F.3:** Let $U_1, U_2, \ldots, U_n$ be independent $\mathsf{Uni}[0,1]$ random variables and $n \geq 1$.

(a) **(10 pts)** Let $U_{(n)} = \max_{i \leq n} U_i$ be the maximum of the $U_i$. Give the density $f_n$ of $U_{(n)}$.

(b) **(5 pts)** Give $\mathbb{E}[U_{(n)}]$.

**Question F.4:** Consider predicting a Bernoulli random variable $Y \in \{0, 1\}$ using a measurement $X$. ($X$ is a random variable.) Suppose you are given the *log odds*

$$f(x) = \log \frac{P(Y = 1 \mid X = x)}{P(Y = 0 \mid X = x)} = \log \frac{P(Y = 1 \mid X = x)}{1 - P(Y = 1 \mid X = x)}.$$

(a) **(7pts)** Give $P(Y = 1 \mid X = x)$ as a function of $f(x)$.

(b) **(7pts)** Say we wish to find the best predictor $\widehat{p}(x)$ of $Y$ given $X = x$ in the mean-square sense, that is, the $\widehat{p}$ solving

$$\underset{g}{\text{minimize}} \ \mathbb{E}\left[(Y - g(X))^2\right]$$

across all functions $g$. What is $\widehat{p}$?

**Question F.5** (Some true/false questions)**:** Let $X$ and $Y$ be random variables, and let $X \perp Y$ denote that $X$ and $Y$ are independent. You do not need to justify your answers.

(a) **(4pts)** If $X$ is constant, then $X \perp Y$.

(b) **(4pts)** Whenever $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$, it must be the case that $X \perp Y$.

(c) **(4pts)** If $\mathrm{Cov}(X, Y) = 0$, then $X \perp Y$.

(d) **(4pts)** If $\mathrm{Cov}(X, Y) > 0$, then $X$ and $Y$ are *not* independent.

(e) **(4pts)** If $\mathrm{Cov}(X, Y)/\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)} > 1$, then something has gone deeply wrong with a calculation somewhere.

**Question F.6 (10pts):** John, obsessed as he is with coffee, claims he can taste the difference between an Americano (in which hot water is poured on top of espresso) and a long black (in which an espresso is poured into a cup of hot water). The TAs have designed a randomized experiment in which they give John 8 cups, four of which are Americanos and four of which are long blacks, and ask him to choose the 4 Americanos. John selects out the 4 Americanos correctly.

The TAs originally believed John was just as likely to select any cup of the coffees. What is the probability that John got all 4 correct by chance? Do you think it was really just chance?

**Question F.7 (12pts):** In a particular population of patients, a proportion $p = \frac{1}{2}$ of the patients have kidney disease. Let $Z$ denote the level of LDL cholesterol above 100 milligrams per deciliter in a given measured patient. In patients with kidney disease, we know that $Z \sim \mathsf{Exp}(1/5)$, while in patients without kidney disease, we know that $Z \sim \mathsf{Exp}(1)$.

For a particular patient, we observe cholesterol level $Z = 5$. What is the probability that the patient has kidney disease given this cholesterol level?

**Question F.8** (Rare events)**:** Two engineers wish to build a sea wall to protect a town from waves. The wall should be built so that (with high probability) it is taller than the tallest wave that will hit the coastline in the next several decades. The first engineer believes that the highest wave height (in meters) $X$ on a given day follows an $\mathsf{Exp}(1)$ distribution, that is, $X$ has density

$$f_1(x) = e^{-x} \quad \text{for } x \geq 0.$$

The second engineer says "To protect against rare events, we should assume the distribution of the waves has heavier tails!" They assert the highest wave height $Z$ has density

$$f_2(z) = \frac{2}{(z+1)^3} \quad \text{for } z \geq 0.$$

(a) **(4pts)** Assuming $X$ has density $f_1$, give $\mathbb{E}[X]$.

(b) **(6pts)** Assuming $Z$ has density $f_2$, give $\mathbb{E}[Z]$. *Hint.* Compute $\mathbb{E}[Z+1]$, then use linearity of expectation.

The engineers now decide to choose a height $h$ so that with probability at least $1-\alpha$, the highest wave across *all* days $1, 2, \ldots, n$, where $n$ is large, are below the height of the wall.

(c) **(4pts)** For $X$ with density $f_1(x) = e^{-x}$ for $x \geq 0$, give $P(X \geq h)$.

(d) **(6pts)** For $Z$ with density $f_2(z) = \frac{2}{(z+1)^3}$ for $z \geq 0$, give $P(Z \geq h)$.

Now we calculate the heights necessary to protect the town. These parts are challenging and the point values do not reflect the difficulty.

(e) **(5pts)** Assume the wave heights $X_1, X_2, \ldots, X_n$ are independent $\mathsf{Exp}(1)$ random variables. Show that if

$$h_n = \log \frac{n}{\alpha}$$

then

$$\lim_{n \to \infty} P\left(\max_{i \leq n} X_i \geq h_n\right) = 1 - e^{-\alpha}.$$

*Hint.* It may be useful to recall that $(1 - c/n)^n \to e^{-c}$ as $n \to \infty$.

*Remark.* As $e^{-\alpha} \geq 1 - \alpha$ and $(1 - \alpha/n)^n$ increases in $n$, for engineer 1, the height $h_n = \log \frac{n}{\alpha}$ guarantees that with probability at least $1 - \alpha$, no wave will surmount the wall in $n$ days.

(f) **(5pts)** Instead suppose the wave heights $Z_1, Z_2, \dots, Z_n$ are independent and have density $f_2$. We wish the wall to be taller than each wave for $n = 365$ days with probability of failure $\alpha = ne^{-10} \approx .016571 = 1.6571\%$. Show that with the choice of $h_n$ in part (e),

$$P\left(\max_{i \le n} Z_i \ge h_n\right) \ge 1 - \exp\left(-\frac{365}{121}\right) \approx .951 = 95.1\%.$$

*Hint.* For all $c \in \mathbb{R}$, we have $e^{-c} \ge 1 - c$.

**Question F.9**$^\star$ (Extra credit, **5pts**): Let $X_1, X_2, X_3, \ldots$ be random variables, sequentially constructed via

$$X_{k+1} = aX_k + Z_{k+1},$$

where $X_0 = 0$ and $Z_k$ are independent $\mathsf{N}(0, 1)$ random variables. Let $P_k$ denote the distribution of $X_k$. Assuming $|a| < 1$, what is the limit $P_* = \lim_{k\to\infty} P_k$?

Scratch Page

Scratch Page