Sample Problems—Advanced Probability

1 Suppose that F is a d.f. and $F(x) = \int_{-\infty}^{x} f(t)dt$ for any x with a continuous f. Then $F' = f \geq 0$ everywhere.

Answer: For an arbitrary x, we show that $F'(x) = \lim_{\Delta x \to 0^+}(F(x+\Delta x) - F(x))/\Delta x$ is in existence and equal to $f(x)$. Since f is continuous, we have for any $\epsilon > 0$ that $|f(x') - f(x)| < \epsilon$ when $|x - x'| < \delta$ for some $\delta > 0$. Thus, due to the nature of Lebesgue integration, as long as $0 < |\Delta x| < \delta$.

$$f(x) - \epsilon < \frac{F(x + \Delta x) - F(x)}{\Delta x} = \frac{\int_{x}^{x+\Delta x} f(t)dt}{\Delta x} < f(x) + \epsilon.$$

In the above, $\int_{x}^{x+\Delta x} f(t)dt/\Delta x$ would mean $\int_{x+\Delta x}^{x} f(t)dt/(-\Delta x)$ when $\Delta x < 0$. This shows that $F'(x)$ is well defined and equal to $f(x)$. Finally, $F'(x) \geq 0$ because as a d.f., F is increasing.

2 The collection $\mathscr{T}$ of all subsets of Ω is a B.F. called the total B.F. If Ω has exactly n points, show that $\mathscr{T}$ has 2^n members.

Answer: Any element ω in Ω is either selected or not selected in a subset A of $\mathscr{T}$. Thus, each element A of $\mathscr{T}$ can be denoted as an n-bit string; for instance, $\emptyset$ corresponds to the all-0 string and Ω the all-1 string. There are 2^n such strings.

3 If Ω is countable, then the total B.F. $\mathscr{T}$ is the same as $\mathscr{F}$ that is the minimal B.F. containing all the singletons.

Answer: For (a), since $\mathscr{F} \subseteq \mathscr{T}$, we just need to show $\mathscr{T} \subseteq \mathscr{F}$. For any $A \in \mathscr{T}$ and equivalently $A \in \Omega$, note A is countable since Ω is. So $A = \{\omega_1, \omega_2, ...\}$ for some sequence $(\omega_i)_{i=1,2,...}$ inside Ω. But by $\mathscr{F}$'s closure under countable unions and the fact that each $\{\omega_i\}$ is an element of $\mathscr{F}$, we have $A = \bigcup_{i=1}^{+\infty}\{\omega_i\} \in \mathscr{F}$ as well.

4 Let $\mathscr{N}$ be the collection of sets that are subsets of null sets, and let $\overline{\mathscr{F}}$ be the collection of subsets of Ω each of which differs from a set in $\mathscr{F}$ by a set in $\mathscr{N}$:

$$\overline{\mathscr{F}} = \{E \subseteq \Omega : E \triangle F \in \mathscr{N} \text{ for some } F \in \mathscr{F}\}.$$

Show that $\overline{\mathscr{F}}$ is also the collection of sets of the form $F \cup N$.

Answer: For any $E \subseteq \Omega$ such that $E \triangle F \in \mathscr{N}$ for some $F \in \mathscr{F}$, let N be the null set that contains $E\triangle F$. For convenience, let $N' = E\setminus F$, $F' = E\cap F$, $C = F\setminus E$, and $D = N\setminus(E\triangle F)$. Note the four sets are disjoint, $F = F'\cup C \in \mathscr{F}$, and $N = N'\cup C\cup D \in \mathscr{F}$. Thus, $F' = (F'\cup C)\setminus(N'\cup C\cup D)$ is in $\mathscr{F}$ as well. But $E = F'\cup N'$ with $F'\in\mathscr{F}$ and $N'\subseteq N$.

On the flip side, suppose $E = F\cup N$ for some $F\in\mathscr{F}$ and $N\in\mathscr{N}$. Then $E\triangle F = N$ which is a member of $\mathscr{N}$.

5 If two r.v.'s are equal a.e., then they have the same p.m.

Answer: When random variables X and Y are equal a.e., we have $\mathbb{P}[X\neq Y] = 0$. Now for any $B\in\mathscr{B}^1$, we have $\mathbb{P}[X\in B] = \mathbb{P}[X\in B \text{ and } X = Y] + \mathbb{P}[X\in B \text{ and } X\neq Y] \leq \mathbb{P}[Y\in B] + \mathbb{P}[X\neq Y] = \mathbb{P}[Y\in B]$. But symmetrically, $\mathbb{P}[Y\in B]\leq\mathbb{P}[X\in B]$ as well. Thus, X and Y must have the same probability measure.

6 If f is Borel measurable, and X and Y are identically distributed, then so are $f(X)$ and $f(Y)$.

Answer: For any $B\in\mathscr{B}^1$, we have

$$\mathbb{P}[f(X)\in B] = \mathbb{P}[X\in f^{-1}(B)] = \mathbb{P}[Y\in f^{-1}(B)] = \mathbb{P}[f(Y)\in B].$$

So $f(X)$ and $f(Y)$ are sharing the same distribution.

7 Let $X\geq 0$ and $\int_\Omega X d\mathbb{P} = A$ for some $A\in(0,+\infty)$. Then the set function ν defined on $\mathscr{F}$ through $\nu(\Lambda) = \int_\Lambda X d\mathbb{P}/A$ is a probability measure on $\mathscr{F}$.

Answer: First, $\nu(\Lambda)\geq 0$ for any $\Lambda\in\mathscr{F}$. Second, let $\Lambda_1,\Lambda_2,\ldots$ be disjoint subsets in $\mathscr{F}$. Then, 证明probability measure 的三个性质

$$\nu\left(\bigcup_{n=1}^{+\infty}\Lambda_n\right) = \frac{1}{A}\int_{\bigcup_{n=1}^{+\infty}\Lambda_n} X d\mathbb{P} = \frac{1}{A}\sum_{n=1}^{+\infty}\int_{\Lambda_n} X d\mathbb{P} = \sum_{n=1}^{+\infty}\nu(\Lambda_n).$$

Lastly, $\nu(\Omega) = \int_\Omega X d\mathbb{P}/A = A/A = 1$.

8 Prove that if $0\leq r < r'$ and $\mathbb{E}[|X|^{r'}] < +\infty$, then $\mathbb{E}[|X|^r] < +\infty$.

Answer: By Hölder's inequality with $p = r'/r$,

$$\mathbb{E}[|X|^r] \leq (E[(|X|^r)^{r'/r}])^{r/r'} = (\mathbb{E}[|X|^{r'}])^{r/r'}.$$

Therefore, $\mathbb{E}[|X|^r]$ would be finite when $\mathbb{E}[|X|^{r'}]$ is.

9 Suppose that $\nu(y:(x,y)\in E)=\nu(y:(x,y)\in F)$ for all x. Show that $(\mu\times\nu)(E)=(\mu\times\nu)(F)$.

Answer: By the definition of the product measure $\mu\times\nu$,

$$(\mu\times\nu)(E)=\int_X \nu(y:(x,y)\in E)\mu(dx)=\int_X \nu(y:(x,y)\in F)\mu(dx)=(\mu\times\nu)(F).$$

10 If $0\leq X_n$, $X_n\leq X\in\mathcal{L}^1$ and $X_n\to X$ in probability, then $X_n\to X$ in $\mathcal{L}^1$.

Answer: Since $X_n\leq X$, we need to show $\mathbb{E}[X-X_n]\to 0^+$. Now for any $\epsilon>0$,

$$\mathbb{E}[X-X_n]=\int_{X-X_n\leq\epsilon}(X-X_n)d\mathbb{P}+\int_{X-X_n>\epsilon}(X-X_n)d\mathbb{P}\leq\epsilon+\int_{X-X_n>\epsilon}X d\mathbb{P},$$

where the inequality takes advantage of $X_n\geq 0$. But $\mathbb{P}[X-X_n>\epsilon]\to 0^+$ would lead to the convergence of the second term to 0^+ as well; see Exercise 2 of Section 3.2. Due to the arbitrariness of ϵ, we therefore have $\mathbb{E}[X-X_n]\to 0^+$.

11 Convergence in $\mathcal{L}^p$ implies that in $\mathcal{L}^r$ for $r<p$.

Answer: By Hölder's inequality,

$$\mathbb{E}[|X_n|^r]\leq(\mathbb{E}[|X_n|^p])^{r/p}.$$

12 Suppose that for any $a<b$,

$$\mathbb{P}[X_n<a\text{ i.o. and }X_n>b\text{ i.o.}]=0.$$

Then, $\lim_{n\to+\infty}X_n$ exists a.e. but it may be infinite.

Answer: Denote the set mentioned in the problem by $A(a,b)$. The hypothesis is that $\mathbb{P}[A(a,b)]=0$ for any $a<b$. Now let B be the set of ω's at which $X_n(\omega)$ does not converge. Also, let $(a_n,b_n)_{n=1,2,\ldots}$ be the sequence of rational-number pairs satisfying $a_n<b_n$. The key observation, due to the denseness of rational numbers in $\Re$, is

$$B=\bigcup_{n=1}^{+\infty}A(a_n,b_n).$$

Therefore, $\mathbb{P}[B] \leq \sum_{n=1}^{+\infty} \mathbb{P}[A(a_n, b_n)] = 0$.

13 Suppose that $X_n \Rightarrow X$ and that h_n and h are Borel functions. Let E be the set of x for which $h_n(x_n) \to h(x)$ fails for some sequence $x_n \to x$. Suppose $E \in \mathscr{B}^1$ and $\mathbb{P}[X \in E] = 0$. Show that $h_n(X_n) \Rightarrow h(X)$.

Answer: Consider random variables Y_n and Y defined on the same probability space $(\Omega, \mathscr{F}, \mathbb{Q})$ such that each Y_n has the same distribution as X_n, Y has the same distribution as X, and $Y_n(\omega) \to Y(\omega)$ at each $\omega \in \Omega$. When $Y(\omega) \notin E \in \mathscr{B}^1$, we have $h_n(Y_n(\omega)) \to h(Y(\omega))$. But $\mathbb{Q}[Y \in E] = \mathbb{P}[X \in E] = 0$. So $h_n(Y_n) \to h(Y)$ a.e. and hence by Theorem 25.2 of Billingsley (1995), $h_n(Y_n) \Rightarrow h(Y)$. Due to the common distributions, we have $h_n(X_n) \Rightarrow h(X)$.

14 If F_1 and F_2 are d.f.'s such that $F_1 = \sum_j b_j \delta_{a_j}$ and F_2 has density p, show that $F_1 * F_2$ has a density and find it.

Answer: Note that

$$\begin{aligned}(F_1 * F_2)(x) &= \int_{-\infty}^{+\infty} F_1(x-y)dF_2(y) = \int_{-\infty}^{+\infty} \sum_j b_j \mathbf{1}(x-y \geq a_j)p(y)dy \\ &= \int_{-\infty}^{+\infty} \sum_j b_j \mathbf{1}(y+a_j \leq x)p(y)dy = \int_{-\infty}^{x} \sum_j b_j p(y-a_j)dy.\end{aligned}$$

Therefore, $(F_1 * F_2)(x)$ has a density function $\sum_j b_j p(x-a_j)$.

15 The convolution of two discrete distributions with exactly m and n atoms, respectively, has at least $m+n-1$ and at most mn atoms.

Answer: Let $x_1 < x_2 < \cdots < x_m$ be the atoms of F and $y_1 < y_2 < \cdots < y_n$ the atoms of G. Note $F * G$ is the distribution of the sum of two independent random variables say X and Y such that X has distribution F and Y distribution G. Thus, the atoms of $F * G$ must come from the set $\{x_i + y_j\}_{i=1,2,\ldots,m, j=1,2,\ldots,n}$. However,

$$x_1+y_1 < x_1+y_2, x_2+y_1 < x_2+y_2 < x_2+y_3, x_3+y_2 < \cdots < x_{m-1}+y_n, x_m+y_{n-1} < x_m+y_n.$$

So the number of atoms is at least $m+n-1$. The number is certainly at most mn.

16 Show that the independence of X and Y implies that $\mathbb{E}[Y|X] = E[Y]$.

Answer: By definition, $\mathbb{E}[Y|X]$ is $\sigma(X)$-measurable and for any $G \in \sigma(X)$,

$$\int_G \mathbb{E}[Y|X]d\mathbb{P} = \int_G Y d\mathbb{P}.$$

The right-hand side is $\mathbb{E}[Y\mathbf{1}_G]$ which due to the independence between X and Y is equal to $\mathbb{E}[\mathbf{1}_G]\mathbb{E}[Y] = \mathbb{P}[G]\mathbb{E}[Y] = \int_G \mathbb{E}[Y]d\mathbb{P}$. Hence, $\mathbb{E}[Y|X] = \mathbb{E}[Y]$ a.e.

17 Prove for bounded X and Y that $\mathbb{E}[Y\mathbb{E}[X|\mathscr{G}]] = \mathbb{E}[X\mathbb{E}[Y|\mathscr{G}]]$.

Answer: Since X and Y are bounded, we always have integrability. As $\mathbb{E}[X|\mathscr{G}]$ is $\mathscr{G}$-measurable, we have by Theorem 34.3 of Billingsley (1995) that $\mathbb{E}[Y\mathbb{E}[X|\mathscr{G}]|\mathscr{G}] = \mathscr{E}[X|\mathscr{G}]\mathbb{E}[Y|\mathscr{G}]$. By the same token, the right-hand side also equals $\mathbb{E}[X\mathbb{E}[Y|\mathscr{G}]|\mathscr{G}]$. Integrating over the entire Ω, we have

$$\begin{aligned}\mathbb{E}[Y\mathbb{E}[X|\mathscr{G}]] = \mathbb{E}[\mathbb{E}[Y\mathbb{E}[X|\mathscr{G}]|\mathscr{G}]] &= \mathbb{E}[\mathbb{E}[X|\mathscr{G}]\mathbb{E}[Y|\mathscr{G}]] \\ &= \mathbb{E}[\mathbb{E}[X\mathbb{E}[Y|\mathscr{G}]|\mathscr{G}]] = \mathbb{E}[X\mathbb{E}[Y|\mathscr{G}]].\end{aligned}$$

18 Show that, if $\mathscr{H} \subseteq \mathscr{G}$ and $\mathbb{E}[X^2] < +\infty$, then

$$\mathbb{E}[(X - \mathbb{E}[X|\mathscr{G}])^2] \leq \mathbb{E}[(X - \mathbb{E}[X|\mathscr{H}])^2].$$

The dispersion of X about its conditional mean becomes smaller as the B.F. grows.

Answer: We would achieve the desired inequality if we could show

$$\mathbb{E}[(X - \mathbb{E}[X|\mathscr{G}])^2|\mathscr{G}] \leq \mathbb{E}[(X - \mathbb{E}[X|\mathscr{H}])^2|\mathscr{G}].$$

But the left-hand side is equal to

$$\mathbb{E}[X^2|\mathscr{G}] - 2\mathbb{E}[X\mathbb{E}[X|\mathscr{G}]|\mathscr{G}] + (\mathbb{E}[X|\mathscr{G}])^2,$$

which, due to Theorem 34.3 of Billingsley (1995), is equal to $\mathbb{E}[X^2|\mathscr{G}] - (\mathbb{E}[X|\mathscr{G}])^2$. For the same reason, the right-hand side is equal to $\mathbb{E}[X^2|\mathscr{G}] - 2\mathbb{E}[X|\mathscr{G}]\mathscr{E}[X|\mathscr{H}] + (\mathbb{E}[X|\mathscr{H}])^2$. Just because

$$(\mathbb{E}[X|\mathscr{G}])^2 + (\mathbb{E}[X|\mathscr{H}])^2 \geq 2\mathbb{E}[X|\mathscr{G}]\mathbb{E}[X|\mathscr{H}],$$

we have the left-hand side being less than the right-hand side.

19 Define $\mathrm{Var}[X|\mathscr{G}] = \mathbb{E}[(X - \mathbb{E}[X|\mathscr{G}])^2|\mathscr{G}]$. Prove that

$$\mathrm{Var}[X] = \mathbb{E}[\mathrm{Var}[X|\mathscr{G}]] + \mathrm{Var}[\mathbb{E}[X|\mathscr{G}]].$$

Answer: Again by Theorem 34.3 of Billingsley (1995),

$$\mathbb{E}[X\mathbb{E}[X|\mathscr{G}]] = \mathbb{E}[\mathbb{E}[X\mathbb{E}[X|\mathscr{G}]|\mathscr{G}]] = \mathbb{E}[\mathbb{E}[(\mathbb{E}[X|\mathscr{G}])^2|\mathscr{G}]] = \mathbb{E}[(\mathbb{E}[X|\mathscr{G}])^2].$$

Now we have

$$\begin{aligned}\mathbb{E}[\mathrm{Var}[X|\mathscr{G}]] = \mathbb{E}[\mathbb{E}[(X - \mathbb{E}[X|\mathscr{G}])^2|\mathscr{G}]] &= \mathbb{E}[(X - \mathbb{E}[X|\mathscr{G}])^2] \\ &= \mathbb{E}[X^2] - 2\mathbb{E}[X\mathbb{E}[X|\mathscr{G}]] + \mathbb{E}[(\mathbb{E}[X|\mathscr{G}])^2],\end{aligned}$$

which according to the just established equation is $\mathbb{E}[X^2]-\mathbb{E}[(\mathbb{E}[X|\mathscr{G}])^2]$. Meanwhile,

$$\begin{aligned}\mathrm{Var}[\mathbb{E}[X|\mathscr{G}]] = \mathbb{E}[(\mathbb{E}[X|\mathscr{G}] - \mathbb{E}[\mathbb{E}[X|\mathscr{G}]])^2] &= \mathbb{E}[(\mathbb{E}[X|\mathscr{G}] - \mathbb{E}[X])^2 \\ &= \mathbb{E}[(\mathbb{E}[X|\mathscr{G}])^2] - 2\mathbb{E}[\mathbb{E}[X]\mathbb{E}[X|\mathscr{G}]] + \mathbb{E}[(\mathbb{E}[X])^2],\end{aligned}$$

which is merely $\mathbb{E}[(\mathbb{E}[X|\mathscr{G}])^2] - (\mathbb{E}[X])^2$. Therefore,

$$\mathrm{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \mathbb{E}[\mathrm{Var}[X|\mathscr{G}]] + \mathrm{Var}[\mathbb{E}[X|\mathscr{G}]].$$

20 A fair coin is flipped until the first head occurs. Let X denote the number of heads required. Find the entropy $H(X)$ in bits.

Answer: For $n = 1, 2, ...,$ note $p(n) \equiv \mathbb{P}[X = n] = (1/2)^n$. Now

$$H(X) = \sum_{n=1}^{+\infty} p(n)\log\left(\frac{1}{p(n)}\right) = \sum_{n=1}^{+\infty} \frac{n}{2^n} = 2.$$

21 Let $\mathcal{X} \equiv \{1, 2, ..., m\}$. Show that the number of sequences $\mathbf{x} \in \mathcal{X}^n$ satisfying $\sum_{i=1}^n g(x_i)/n \geq \alpha$ is approximately equal to 2^{nH^*}, to the first order in the exponent, for n sufficiently large, where

$$H^* = \max_{P:\sum_{i=1}^m P(i)g(i)\geq\alpha} H(P).$$

Answer: Let Q the uniform distribution on $\mathcal{X}$. Note

$$D(P||Q) = \sum_{i=1}^{m} P(i)\log\left(\frac{P(i)}{Q(i)}\right) = \log(m) - H(P).$$

For $E = \{P : \sum_{i=1}^{m} P(i)g(i) \geq \alpha\}$, the P^* that maximizes $H(P)$ for $P \in E$ is the same as the one that minimizes $D(P||Q)$ for $P \in E$. By Sanov's theorem, we have $Q^n(E) \doteq 2^{-nD(P^*||Q)} = 2^{nH(P^*)}/m^n$ when n is large.

At the same time, note that $\mathbf{x} \in \mathcal{X}^n$ satisfying $\sum_{i=1}^{n} g(x_i)/n \geq \alpha$ if and only if $P_{\mathbf{x}} \in E$. So the count is just $m^n Q^n(E)$ which is close to $2^{nH(P^*)}$.

22 Suppose that an atom is equally likely to be in each of six states, $X \in \{s_1, s_2, ..., s_6\}$. One observes the states $X_1, X_2, ..., X_n$ of n atoms independently drawn according to this uniform distribution. It is observed that the frequency of occurrence of state s_1 is twice the frequency of occurrence of state s_2.

(a) To first order in the exponent, what is the probability of observing this?

(b) Assuming n large, find the conditional distribution of the state of the first atom X_1, given this observation.

Answer: Let $E = \{(P(1), ..., P(6)) \in [0,1]^6 : P(1) = 2P(2), P(1) + \cdots + P(6) = 1\}$ and $Q(1) = Q(2) = \cdots = Q(6) = 1/6$. Note

$$D(P||Q) = \sum_{i=1}^{6} P(i) \log\left(\frac{P(i)}{Q(i)}\right) = \log(6) + \sum_{i=1}^{6} P(i)\log(P(i)).$$

The problem of $\min_{P\in E} D(P||Q)$ can be solved through the following:

$$\begin{array}{ll} \min & \sum_{i=1}^{6} P(i)\log(P(i)) \\ \text{s.t} & P(1) - 2P(2) = 0 \\ & P(1) + P(2) + \cdots + P(6) = 1 \\ & P(1), P(2), ..., P(6) \geq 0. \end{array}$$

Note this is a convex program. So a feasible $(P(1), ..., (P(6))$ will be optimal if and only if there exist α, β, and $\gamma(1) \geq 0, ..., \gamma(6) \geq 0$, so that $\log(P(1)) + 1 - \alpha - \beta - \gamma(1) = 0$, $\log(P(2)) + 1 + 2\alpha - \beta - \gamma(2) = 0$, $\log(P(3)) + 1 - \beta - \gamma(3) = 0$, $\cdots$, $\log(P(6)) + 1 - \beta - \gamma(6) = 0$, and $\gamma(1)P(1) = \cdots = \gamma(6)P(6) = 0$. We can get

$$P(1) = \frac{2^{1/3}}{2^{1/3} + 2^{-2/3} + 4}, \qquad P(2) = \frac{2^{-2/3}}{2^{1/3} + 2^{-2/3} + 4},$$

$$P(3) = P(4) = P(5) = P(6) = \frac{1}{2^{1/3} + 2^{-2/3} + 4},$$

$\alpha = 1/3$, $\beta = \log(2/(2^{1/3} + 2^{-2/3} + 4))$, and $\gamma(1) = \cdots = \gamma(6) = 0$.

(a) To first order in the exponent, the chance to observe this event is $e^{-nD(P||Q)}$ where the $P(i)$'s provided as in the above and the $Q(i)$'s are all equal to $1/6$.

(b) The conditional distribution of the state of the first atom will be $P = (P(1), ..., P(6))$ as provided in the above.