Statistics 330/600, spring 2018
Homework #9 solutions

*[1] *Suppose $(\mathcal{X}, \mathcal{A}, \mathbb{P})$ is a probability space and $T : \mathcal{X} \to \mathcal{Y} = \{1, 2, \ldots, k\}$ is a map for which $B_i = \{x : Tx = i\}$ is $\mathcal{A}$-measurable, for each i. Let $\mathcal{B}$ denote the sigma-field generated by the sets $B_1, \ldots, B_k$.*

Solution: *Implicitly the set $\mathcal{Y}$ is equipped with the sigma-field $\mathcal{C}$ consisting of all subsets and T is $\mathcal{B}\backslash\mathcal{C}$-measurable because*

$$T^{-1}(C) = \cup_{i \in C} B_i \in \mathcal{B} \qquad \text{for each } C \text{ in } \mathcal{C}.$$

The distribution of T under $\mathbb{P}$ is the probability measure $\mathbb{Q}$ on $\mathcal{C}$ for which $\mathbb{Q}\{i\} = \mathbb{P}(T^{-1}\{i\}) = \mathbb{P}B_i$.

Many of you got the functions g and T confused.

(i) Show that each g in $\mathcal{M}^+(\mathcal{X}, \mathcal{B})$ must be of the form $\sum_{i=1}^k \alpha_i \mathbb{1}\{x \in B_i\}$, for constants α_i with $0 \le \alpha_i \le \infty$. Hint: Why could there not exist points x_1 and x_2 in some B_i for which $g(x_1) \ne g(x_2)$?

Solution: *If x_1 and x_2 belong to the same B_i then $\mathbb{1}_E(x_1) = \mathbb{1}_E(x_2)$ for each E in the generating class $\mathcal{E} = \{B_1, \ldots, B_k\}$. You can check that*

$$\mathcal{B}_0 = \{B \in \mathcal{B} : \mathbb{1}_B(x_1) = \mathbb{1}_B(x_2)\}$$

is a sigma-field for which $\mathcal{B}_0 \supset \mathcal{E}$. It follows that $\mathcal{B}_0 = \mathcal{B}$. For every B in $\mathcal{B}$, either $\{x_1, x_2\} \subset B$ or $\{x_1, x_2\} \subset B^c$. Equivalently, if $B \in \mathcal{B}$ and $B \cap B_i \ne \emptyset$ then $B \supset B_i$. Again equivalently, if B_i is partitioned into the union of two disjoint sets in $\mathcal{B}$ then one of the sets must be empty.

Remark. *The B_i's are called **atoms** of the sigma-field $\mathcal{B}$ because none of them can be split into a non-trivial union of two disjoint $\mathcal{B}$-sets. Another way to prove that the B_i's are atoms is to show that every non-empty member of $\mathcal{B}$ is a finite union of B_i's.*

If $g(x_1) = a_1 \ne a_2 = g(x_2)$ then $B := g^{-1}\{a_1\} \cap B_i \in \mathcal{B}$ and $\mathbb{1}_B(x_1) = 1 \ne 0 = \mathbb{1}_B(x_2)$: a contradiction. We cannot use a $\mathcal{B}$-measurable function to split the atoms of $\mathcal{B}$. Thus g must take a constant value, say α_i, on each set B_i, that is, $g(x) = \sum_{i=1}^k \alpha_i \mathbb{1}\{x \in B_i\}$. The $\mathcal{B}\backslash\mathcal{B}[0, \infty]$-measurable functions must be constant on each atom of $\mathcal{B}$.

(ii) If $\mathbb{P}B_i > 0$ define $K_i f = \mathbb{P}(f(x)\mathbb{1}\{x \in B_i\})/\mathbb{P}B_i$ for $f \in \mathcal{M}^+(\mathcal{X}, \mathcal{A})$. If $\mathbb{P}B_i = 0$ let K_i be the zero measure on $\mathcal{A}$, that is, $K_i f = 0$ for $f \in$

$\mathcal{M}^+(\mathcal{X}, \mathcal{A})$. *Show that* $\mathbb{K} = \{K_i : 1 \le i \le k\}$ *provides a conditional distribution for* $\mathbb{P}$ *given* T.

SOLUTION: *By construction*

$$K_i\{x : Tx \neq i\} = K_i B_i^c = \begin{cases} 0 & \text{if } \mathbb{P}B_i = 0 \\ \mathbb{P}(B_i^c B_i)/\mathbb{P}B_i = 0 & \text{if } \mathbb{P}B_i \neq 0. \end{cases}$$

Remark. *Actually I didn't need to worry about those* i *for which* $\mathbb{Q}\{i\} = 0$ *because we only need* $K_i\{x : Tx \neq i\} = 0$ a.e. $[\mathbb{Q}]$ *and* $K_i\mathcal{X} = 1$ a.e. $[\mathbb{Q}]$.

Every function from $\mathcal{Y}$ *into* $\overline{\mathbb{R}}$ *is* $\mathcal{C}$*-measurable because every subset of* $\mathcal{Y}$ *belongs to* $\mathcal{C}$. *In particular,* $i \mapsto K_i f$ *is* $\mathcal{C}$*-measurable.*

To show that $\mathbb{K}$ *plays the role of a conditional distribution we also need to show that*

$$\mathbb{P}f = \mathbb{Q}^y K_y^x f(x) = \sum\nolimits_{i \in \mathcal{Y}} \mathbb{Q}\{i\} K_i f \qquad \text{for each } f \text{ in } \mathcal{M}^+(\mathcal{X}, \mathcal{F}).$$

This follows by linearity and the fact that $\sum_{i \in \mathcal{Y}} \mathbb{1}_{B_i} = 1$*:*

$$\mathbb{P}f = \sum\nolimits_{i \in \mathcal{Y}} \mathbb{P}(f \mathbb{1}_{B_i}) = \sum\nolimits_{i \in \mathcal{Y}} (\mathbb{P}B_i)(K_i f).$$

The moral: In the case where T *takes on only finitely (or countably) many different values the conditional distributions can be defined (in elementary fashion) by taking ratios.*

(iii) Find $\mathbb{P}_{\mathcal{B}} f$ *for* $f \in \mathcal{M}^+(\mathcal{X}, \mathcal{A})$.

SOLUTION: *If you could fight your way through all the abstractions you might remember that* $\mathbb{P}_{\mathcal{B}} f$ *is just a function* F *in* $\mathcal{M}^+(\mathcal{X}, \mathcal{B})$ *for which*

$$\mathbb{P}(fG) = \mathbb{P}(FG) \qquad \text{for each } G \text{ in } \mathcal{M}^+(\mathcal{X}, \mathcal{B}).$$

The function F *must be of the form* $\sum_{i \in \mathcal{Y}} \beta_i \mathbb{1}\{x \in B_i\}$. *The general* G *is representable as* $\sum_{i \in \mathcal{Y}} \alpha_i \mathbb{1}\{x \in B_i\}$. *Thus we need*

$$\sum\nolimits_{i \in \mathcal{Y}} \alpha_i \beta_i \mathbb{P}B_i = \mathbb{P}(FG) = \mathbb{P}(fG) = \sum\nolimits_{i \in \mathcal{Y}} \alpha_i \mathbb{P}(f \mathbb{1}_{B_i}).$$

By considering cases where only one of the α_i*'s is non-zero you can check that the solution for* F *is given by*

$$\beta_i = \begin{cases} \mathbb{P}(f \mathbb{1}_{B_i})/\mathbb{P}B_i & \textit{if } \mathbb{P}B_i \neq 0 \\ \text{arbitrary} & \textit{if } \mathbb{P}B_i = 0. \end{cases}$$

The only ambiguity in the choice of F *comes from its values on sets* B_i *with* $\mathbb{P}B_i = 0$. *In short,* $F(x) := \sum_{i \in \mathcal{Y}} \mathbb{1}\{x \in B_i\} K_i f$ *is one of* *the $\mathbb{P}$-equivalence class of $\mathcal{M}^+(\mathcal{X}, \mathcal{B})$ functions that can play the role of $(\mathbb{P}_{\mathcal{B}} f)(x)$.*

In class I tried to argue that we don't really need the Kolmogorov abstraction when a conditional probability distribution exists. This problem was my attempt to provide some clarification.

If, as in this problem, a conditional distribution exists then the function $\kappa(y, f) := K_y^x f(x)$, the expected value of f with respect to the conditional distribution, can be taken as the conditional expectation $\mathbb{P}(f \mid T = y)$. The Kolmogorov construction for $\kappa(y, f)$ provides a function with analogous properties to $K_y f$. The construction is most useful as a substitute for 'expectations with respect to conditional distributions' when the full conditional distribution does not exist. The Kolmogorov conditional expectation, $\mathbb{P}(f \mid T)$ or $\mathbb{P}_{\mathcal{B}} f$, as a function on $\mathcal{X}$ is defined as $\kappa(Tx, f)$. It corresponds to writing $\mathbb{Q}^y g(y)\kappa(y, f)$ as $\mathbb{P} g(Tx)\kappa(Tx, f)$.

Note that chaos ensues in the current problem if you write $\kappa(Tx, f)$ as $K_{Tx}^x f(x)$, because the dummy variable of integration, x, is now confused with the argument of the T. It would be a lot like defining a function on $\mathbb{R}$ by

$$H(y) = \int_0^1 (y + x)^2 dx = y^2 + y + 1/3$$

then evaluating H at $y = \cos(x)$ to get $H(\cos x) = \cos^2 x + \cos x + 1/3$. You would get something quite different from the integral

$$\int_0^1 (\cos(x) + x)^2 dx.$$

Instead of a function of x you now have a single real number. It is, however, true that

$$H(\cos x) = \int_0^1 (\cos(x) + z)^2 dz,$$

in the same way that $\kappa(Tx, f) = K_{Tx}^z f(z)$.

*[2] *(Neyman factorization theorem cf. UGMTP Example 5.31) Suppose $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $\{\mathbb{P}_\theta : \theta \in \Theta\}$ is a set of probability measures on $\mathcal{F}$ with each $\mathbb{P}_\theta$ absolutely continuous with respect to $\mathbb{P}$. Suppose also that $\mathcal{G}$ is a sub-sigma-field of $\mathcal{F}$ and that $\mathbb{P}_\theta$ has density*

$$p(\omega, \theta) = g_\theta(\omega) h(\omega) \qquad \text{with } g_\theta \in \mathcal{M}^+(\Omega, \mathcal{G}) \text{ for each } \theta$$

for a fixed $h \in \mathcal{M}^+(\Omega, \mathcal{F})$ that doesn't depend on θ. That is, $\mathbb{P}_\theta f = \mathbb{P}\left(g_\theta(\omega) h(\omega) f(\omega)\right)$ for each f in $\mathcal{M}^+(\Omega, \mathcal{F})$.

Let H be a function (unique up to $\mathbb{P}$-equivalence) in $\mathcal{M}^+(\Omega, \mathcal{G})$ for which

$$\mathbb{P}g(\omega)h(\omega) = \mathbb{P}g(\omega)H(\omega) \qquad \text{for each } g \text{ in } \mathcal{M}^+(\Omega, \mathcal{G}).$$

(That is H is a possible choice for $\mathbb{P}(h \mid \mathcal{G})$.)

SOLUTION: *Remember that*

<1> $$\mathbb{P}_{\mathcal{G}} h = H(\omega) \qquad \text{a.e. } [\mathbb{P}]$$

<2> $$\mathbb{P}_{\mathcal{G}}(Xh) = \gamma(\omega) \qquad \text{a.e. } [\mathbb{P}]$$

That is, both H and γ are in $\mathcal{M}^+(\Omega, \mathcal{G})$ and

<3> $$\mathbb{P}(hg) = \mathbb{P}(Hg)$$

<4> $$\mathbb{P}(Xhg) = \mathbb{P}(\gamma g)$$

for every g in $\mathcal{M}^+(\Omega, \mathcal{G})$.

(i) Show that $\mathbb{P}_\theta\{H = 0\} = 0$ for each θ.

SOLUTION:

$$\begin{aligned}
\mathbb{P}_\theta\{H = 0\} &= \mathbb{P}h(\omega)g_\theta(\omega)\{H(\omega) = 0\} && \text{density} \\
&= \mathbb{P}H(\omega)g_\theta(\omega)\{H(\omega) = 0\} && \text{by <3> with } g = g_\theta\{H(\omega) = 0\} \\
&= 0 \qquad \text{because } H(\omega)\{H(\omega) = 0\} = 0 \text{ for every } \omega.
\end{aligned}$$

(ii) Show that $\mathbb{P}_\theta\{H = \infty\} = 0$ for each θ. Hint: Start by showing that

$$1 \geq n\mathbb{P}g_\theta \mathbb{1}\{H(\omega) = \infty\} \qquad \text{for each } n \in \mathbb{N}.$$

What does that tell you about $g_\theta(\omega)\mathbb{1}\{H(\omega) = \infty\}$?

SOLUTION:

$$\begin{aligned}
1 &\geq \mathbb{P}_\theta\{H = \infty\} \\
&= \mathbb{P}g_\theta h\{H = \infty\} && \text{density} \\
&= \mathbb{P}g_\theta H\{H = \infty\} && \text{by <3> with } g = g_\theta\{H = \infty\} \\
&\geq \mathbb{P}g_\theta n\{H = \infty\} && \text{because } H \geq n \text{ on } \{H = \infty\}.
\end{aligned}$$

Divide both sides of the inequality by n then let n tend to ∞ to deduce that $\mathbb{P}g_\theta\{H = \infty\} = 0$. It follows that $g_\theta\{H = \infty\} = 0$ a.e. $[\mathbb{P}]$ which implies that $g_\theta(\omega)h(\omega)\{H(\omega) = \infty\} = 0$ a.e. $[\mathbb{P}]$. Integrate with respect to $\mathbb{P}$ to get

$$0 = \mathbb{P}g_\theta(\omega)h(\omega)\{H(\omega) = \infty\} = \mathbb{P}_\theta\{H = \infty\}.$$

(iii) For each X in $\mathcal{M}^+(\Omega, \mathcal{F})$ and some fixed choice of γ of $\mathbb{P}(Xh \mid \mathcal{G})$ define

$$Y(\omega) = \frac{\gamma(\omega)}{H(\omega)} \mathbb{1}\{0 < H < \infty\}.$$

Show that $\mathbb{P}_\theta(X \mid \mathcal{G}) = Y$ *a.e* $[\mathbb{P}_\theta]$ *for every* θ.

SOLUTION: *You needed to show that* $Y \in \mathcal{M}^+(\Omega, \mathcal{G})$ *and*

<5> $$\mathbb{P}_\theta(XG) = \mathbb{P}_\theta(YG) \qquad \text{for each } G \text{ in } \mathcal{M}^+(\Omega, \mathcal{G}).$$

The ratio of two $\mathcal{G}$*-measurable functions is* $\mathcal{G}$*-measurable provided there are no* ∞/∞ *or* $0/0$ *problems. For* Y *the indicator function in the definition eliminated those pesky cases. For* <5>*, start from the right-hand side.*

$$\begin{aligned}
\mathbb{P}_\theta(YG) &= \mathbb{P}\left(g_\theta h Y G\right) && \text{density} \\
&= \mathbb{P}(g_\theta H Y G) && \text{by } <3> \\
&= \mathbb{P}(g_\theta \gamma \{0 < H < \infty\} G) && \text{definition of } HY \\
&= \mathbb{P}(g_\theta X h \{0 < H < \infty\} G) && \text{by } <4> \\
&= \mathbb{P}_\theta(X\{0 < H < \infty\}G).
\end{aligned}$$

Parts (i) and (ii) show that the last expression is unchanged if we add the $\mathbb{P}_\theta$*-negligible terms* $XG\{H = 0\}$ *and* $XG\{H = \infty\}$ *to the final integrand.*