UH, Probability Theory Fall 2015, Solutions Problems 6 (14.10.2015)

1. When the cumulative distribution function $F_X(t) = P(X \le t)$ of a $\mathbb{R}$-valued random variable X which is absolutely continuous with respect to Lebesgue measure, which means

$$F_X(b) = F_X(a) + \int_a^b f_X(t)dt$$

for some Borel measurable function $f_X(t) \ge 0$, which is called probability density function. When the classical derivative $\frac{dF_X}{dt}(t)$ exists at all t, then it is a probability density function. More in general $\frac{dF_X}{dt}(t) = \frac{dP_X}{dt}(t) =$ is understood as the Radon Nikodym derivative of the push-forward probability measure P_X with respect to Lebesgue measure.

In such case, for every non-negative and Borel measurable test function $g(x) \ge 0$ we have

$$\begin{aligned} E_P\big(g(X)\big) &= \int_\Omega g(X(\omega))P(d\omega) = \int_{\mathbb{R}} g(t)P_X(dt) \\ &= \int_{\mathbb{R}} g(t)F(dt) = \int_{\mathbb{R}} g(t)f_X(t)dt \end{aligned} \tag{0.1}$$

where $P_X(B) = \mathbb{P}\big(\{\omega : X(\omega) \in B\}\big)$ is the pushforward measure of $\mathbb{P}$ by the random variable X. The integral w.r.t. P_X on $\mathbb{R}$ is the same as the Lebesgue Stieltjes integral w.r.t dF, meaning that P_X coincides with the measure induced by the cumulative distribution function $F(t)$ on $\mathbb{R}$.

Hint: One possible strategy for this proof is to use the monotone class theorem: Define the class

$$\mathcal{C} = \Big\{ g : \mathbb{R} \to [0,\infty) \text{ bounded and Borel measurable such that (0.3) holds } \Big\}$$

and show that $\mathcal{C}$ is a monotone class (use the linearity of the integral together with the monotone convergence theorem) which contains the indicators $\mathbf{1}_{(a,b]}(t)\ \forall a \le b \in \mathbb{R}$.

Solution The assumption of almost everywhere (w.r.t. Lebesgue measure) differentiability of $F_X(t)$ is equivalent to

$$F(b) = F(a) + \int_a^b f_X(t)dt \quad \forall a,b \in \mathbb{R}\ . \tag{0.2}$$

When $g(t)=\mathbf{1}_{(a,b]}(t)$ we have

$$E_P\big(g(X)\big)=P(X\in(a,b])=F(b)-F(a)=\int_{\mathbb{R}}\mathbf{1}_{(a,b]}(t)f_X(t)dt$$

which means that $\mathbf{1}_{(a,b]}\in\mathcal{C}$. If we show that $\mathcal{C}$ is a monotone class, it follows by the monotone class theorem that $\mathcal{C}$ contains all functions which are bounded and measurable w.r.t. the σ-algebra generated by the intervals $\sigma((a,b]:a\le b\in\mathbb{R})=\mathcal{B}(\mathbb{R})$ which is the Borel σ-algebra.

Lets check: the constant $1\in\mathcal{C}$, since

$$1=E_P(1)=F(+\infty)-F(-\infty)=\int_{-\infty}^{+\infty}f_X(t)dt$$

Since all integrals in (0.2) are linear w.r.t the integrand function g, it follows that $\mathcal{C}$ is a vector space.

Let $(g_n(t):n\in\mathbb{N})\subseteq\mathcal{C}$ with $0\le g_n(t)\uparrow g(t)\le K<\infty$. Then

$$\begin{aligned}E_P\big(g_n(X)\big)&=\int_\Omega g(X(\omega))P(d\omega)=\int_{\mathbb{R}}g_n(t)P_X(dt)\\&=\int_{\mathbb{R}}g_n(t)F(dt)=\int_{\mathbb{R}}g_n(t)f_X(t)dt\end{aligned}\tag{0.3}$$

and since the monotone convergence theorem holds for all the integrals in (0.3) it follows that we can take the limit inside the integral and (0.3) holds for g.

When $g(t)\ge 0$ a non-negative Borel measurable function which is not bounded, let $g^{(N)}(t)=g(t)\wedge N$, then (0.3) holds for each $g^{(N)}(t)$ and once again the monotone convergence theorem implies that (0.3) holds for $g(t)$ as well.

2. **Linearity of the expectation** The expectation of a random variable $X(\omega)$ is defined as

$$E_{\mathbb{P}}\big(X\big)=E_{\mathbb{P}}\big(X^+\big)-E_{\mathbb{P}}\big(X^-\big)$$

where $X^+=\max\{X,0\}\ge 0$, $X^-=\max\{-X,0\}\ge 0$ are non-negative random variables, and we have defined first for non-negative random variables

$$E_{\mathbb{P}}(X)=\sup_{Y\in\mathcal{S}F:0\le Y\le X}\big\{E_{\mathbb{P}}(Y)\big\}$$

In this way the expectation is well defined unless

$E_{\mathbb{P}}(X^+) = E_{\mathbb{P}}(X^-) = +\infty$.

In the lectures we have shown (first for simple random variables and then by the monotone convergence theorem) that when $X(\omega) \geq 0$, $Y(\omega) \geq 0$ P-almost surely (outside a $\mathbb{P}$-null set), and $a, b \geq 0$

$$E_{\mathbb{P}}(aX + bY) = aE_{\mathbb{P}}(X) + bE_{\mathbb{P}}(Y) \tag{0.4}$$

Show that linearity holds for any random variables X, Y and $a, b \in \mathbb{R}$ when the expectations on both left and right sides in (0.4) are finite.

Hint: write $(aX + bY)$ using the representations $X = (X^+ - X^-)$, $Y = (Y^+ - Y^-)$, $a = (a^+ - a^-)$, $b = (b^+ - b^-)$, and integrate the positive parts and negative parts separately.

Solution

$$\begin{aligned} aX + bY =&(aX + bY)^+ - (aX + bY)^- = (a^+ - a^-)(X^+ - X^-) + (b^+ - b^-)(Y^+ - Y^-) = \\ &(a^+X^+ + a^-X^- + b^+Y^+ + b^-Y^-) - (a^-X^+ + a^-X^+ + b^-Y^+ + b^-Y^+) \end{aligned}$$

Then

$$\begin{aligned} E(aX + bY) =&E(a^+X^+ + a^-X^- + b^+Y^+ + b^-Y^-) - E(a^-X^+ + a^-X^+ + b^-Y^+ + b^-Y^+) \\ =&a^+E(X^+) + a^-E(X^-) + b^+E(Y^+) + b^-E(Y^-) + \\ &- a^-E(X^+) + a^-E(X^+) + b^-E(Y^+) + b^-E(Y^+) \\ =&(a^+ - a^-)(E(X^+) - E(X^-)) + (b^+ - b^-)(E(Y^+) - E(Y^-)) \\ =&aE(X) + bE(Y) \end{aligned}$$

where we have used linearity for non-negative random variables with non-negative coefficients.

3. Let $U(\omega)$ be uniformly distributed r.v. with values in $[0, 1]$, such that $\mathbb{P}(\{U \in (a, b]\}) = (b - a)$ for $0 \leq a \leq b \leq 1$.

   (a) Show that the powers $U(\omega)^z$, with $z \in \mathbb{Z}$ (the integers) are random variables. **Solution** The map $u \mapsto u^n$ is continuous when $n \in \mathbb{N}\{0, 1, 2 \ldots\}$, and the $U(\omega)^n$ is a random variable since a continuous function composed with a measurable map is measurable. When $z = -n$ and $n \geq 1$, the map $u \mapsto u^{-n}$ is not continuous at $u = 0$, and $0^{-n} = +\infty$. Nevertheless $U^{-n}(\omega)$ is a random variable since $\{\ \omega : U^{-n}(\omega) \leq t\} = \{\omega : U(\omega) \geq t^{-1/n}\} = U^{-1}([t^{-1/n}, +\infty) \in \mathcal{F}$, since U is a random variable.

(b) Compute the moments $E_{\mathbb{P}}(U^z) \in [0, +\infty]$ for $z \in \mathbb{Z}$.
We distinguish 3 cases: $z \in \mathbb{N}$, $z = -1$ and $z \leq -2$.
For $n \geq 0$

$$E_P(U^n) = \int_0^1 u^n du = \frac{1}{n+1}$$

$$E_P(1/U) = \int_0^1 \frac{1}{u} du = \log(1) - \log(0) = +\infty$$

which implies

$$E_P(U^{-n}) = \infty \quad \forall n \in \mathbb{N}$$

by comparison since $0 \leq U^{-1} \leq U^{-n}$ when $n \geq 1$ and $0 \leq U \leq 1$, and by the positivity of the expectation $+\infty = E_P(U^{-1}) \leq E_P(U^{-n})$.

(c) Compute the exponential moments $E_{\mathbb{P}}\big(\exp(tU)\big)$ for $t \in \mathbb{R}$.

$$E_{\mathbb{P}}\big(\exp(tU)\big) = \int_0^1 \exp(tu) du = \frac{e^t - 1}{t}$$

Note that this is continuous at $t = 0$,

$$\lim_{t \to 0} E_{\mathbb{P}}\big(\exp(tU)\big) = \lim_{t \to 0} \frac{e^t - 1}{t} = \lim_{t=0} \frac{e^t}{1} = 1 = E_P(\exp(t_0 U))$$

with $t_0 = 0$, where we used l'Hospital rule. This is also a consequence of the bounded convegence theorem, for any in $t \in [-T, T]$ with $T > 0$, we have $0 \leq \exp(tU(\omega)) \leq \exp(TU(\omega))$ with $E_P(\exp(TU)) < \infty$ which implies that the map $t \mapsto E_P(\exp(tU))$ is continuous.

(d) Compute the trigonometric moments $E_{\mathbb{P}}\big(\cos(2\pi tU)\big)$ and $E_{\mathbb{P}}\big(\sin(2\pi tU)\big)$ for $t \in \mathbb{R}$.
Solution

$$E_{\mathbb{P}}\big(\sin(2\pi tU)\big) = \int_0^1 \sin(2\pi tu) du = \frac{1 - \cos(2\pi t)}{2\pi t}$$

Note that since $|\sin(x)| \leq 1$, it follows by the dominated convergent theorem that the map $t \mapsto E_{\mathbb{P}}\big(\sin(2\pi tU)\big)$ is continous.
Note for example that we have continuity at $t = 0$:

$$\lim_{t\to 0} E_{\mathbb{P}}\big(\sin(2\pi tU)\big) = \lim_{t\to 0} \frac{1-\cos(2\pi t)}{2\pi t} = \lim_{t\to 0} \frac{2\pi \sin(2\pi t)}{2\pi} = 0 = E_{\mathbb{P}}\big(\sin(2\pi t_0 U)\big)$$

with $t_0 = 0$.

$$E_{\mathbb{P}}\big(\cos(2\pi tU)\big) = \int_0^1 \cos(2\pi tu)du = \frac{\sin(2\pi t)}{2\pi t}$$

Note that since $|\cos(x)| \leq 1$, it follows by the dominated convergent theorem that the map $t \mapsto E_{\mathbb{P}}\big(\cos(2\pi tU)\big)$ is continous.
Note for example that we have continuity at $t = 0$:

$$\lim_{t\to 0} E_{\mathbb{P}}\big(\cos(2\pi tU)\big) = \lim_{t\to 0} \frac{\sin(2\pi t)}{2\pi t} = \lim_{t\to 0} \frac{2\pi \cos(2\pi t)}{2\pi} = 1 = E_{\mathbb{P}}\big(\cos(2\pi t_0 U)\big)$$

with $t_0 = 0$.

4. Let $f : [0, T] \to \mathbb{R}^+$ be a non-negative and bounded measurable function.

   We define its upper and lower Riemann-integrals as follows:

$$J^+(f) = \inf\big\{ I(g) : g \geq f, g \text{ takes finitely many values and is piecewise continuous } \big\}$$
$$J^-(f) = \sup\big\{ I(g) : g \leq f, g \text{ takes finitely many values and is piecewise continuous } \big\}$$

where the integral $I(g)$ of a piecewise continuous function g taking finitely many values is the usual finite sum.

Note that on the real line, a piecewise continuous simple function taking finitely many values is piecewise constant, with representation

$$g(x) = \sum_{k=1}^n a_k \mathbf{1}_{E_i}(x), \text{ with } I(g) = \sum_{k=1}^n a_k \text{ length}(E_i)$$

where E_i are intervals. In the construction of Lebesgue integral, the general definition uses Borel sets instead of intervals.

We say that f is Riemann integrable when $J^+(f) = J^-(f)$ which defines the Riemann integral $J(f)$ (it is possible that $J(f) = +\infty$).

(a) Show that when f is Riemann integrable the Riemann integral $J(f)$ coincides with Lebegue integral $I(f)$ defined in the lectures.

Hint We define the Lebesgue integral $I(f)$ of a Borel measurable non-negative function w.r.t. Lebesgue measure as

$$I(f) = \sup\{I(g) : g \le f, \quad g \text{ is measurable and takes finitely many values }\}$$

(b) Show that a non-negative continuous function f is Riemann integrable on the compact set $[0, T]$.

Hint: a continuous function uniformly continuous on compact sets. Note that you can approximate uniformly on compacts a continuous function by **piecewise continuous** simple functions.

Solution A simple piecewise continuous function has representation

$$g(x) = \sum_{k=1}^{n} y_k \mathbf{1}_{E_k}(x)$$

where E_k are intervals.

Since we are integrating with respect to Lebesgue measure, functions which differ on a set of Lebesgue measure zero have the same integral, so we can assume for example that $E_k = (a_k, b_k]$. Denote by $\mathcal{S}C$ the class of simple piecewise continuous functions. Then for such g Riemann and Lebesgue integrals coincide with the Riemann sum

$$J(g) = I(g) = \sum_{k=1}^{n} y_k \lambda(E_k) = \sum_{k=1}^{n} y_k (b_k - a_k)$$

Let now $f(x) \ge 0$ a Borel measurable function, and let $0 \le g'(x) \le f(x) \le g``(x)$ for some $g', g'' \in \mathcal{SC}^+$. Since the Lebesgue integral is a positive operator,

$0 \le J(g') = I(g') \le I(f) \le I(g'\prime) = J(g$

$)$

by taking the supremum over $g' \in \mathcal{S}C^+$, with $0 \le g' \le f$, and infinum over $g' \in \mathcal{SC}^+$, with $f \le g'$, it follows that

$$0 \le J^-(f) \le I(f) \le J^+(f)$$

Therefore when f is Riemann integrable, by definition $J(f) = J^+(f) = J^-(f) = I(f)$, and by the sandwhiching argument the Lebegue integral and Riemann integral coincide when the latter exists. However the Lebesgue integral is more general, we show in the next example a Borel function $f \geq 0$ with well define Lebesgue integral such that $J^-(f) \lneq J^+(f)$ and the Riemann integral does not exists.

(c) Let $f(x) = \mathbf{1}_{\mathbb{Q}}(x)$ where $\mathbb{Q}$ are the rationals.
Show that f is Borel measurable, but is not Riemann integrable on $[0, T]$.
Hint : Show that on a compact interval $[0, T]$ $J^+(f) = T$ and $J^-(f) = 0$.
Solution Since the rationals are dense in $\mathbb{R}$, it follows that the smallest piecewise continuous simple function which is an upper bound for $\mathbf{1}_Q(x)$ is the constant function with value 1, and t the biggest piecewise continuous simple function which is a lower bound for $\mathbf{1}_Q(x)$ is the constant function with value 0. Therefore $J^+(f) = 1 \times \lambda([0, T]) = T$ and $J^-(f) = 0 \times \lambda([0, T]) = 0$, which shows that $\mathbf{1}_{\mathbb{Q}}(x)$ is not Riemann integrable.

(d) For the Lebesgue integral we have

$$I(f) = \int_0^T f(x)dx = \int_0^T \mathbf{1}_Q(x)dx = \sum_{q \in [0,T] \cap \mathbb{Q}} \lambda(\{q\}) = 0$$

since the Lebesgue measure assigns zero mass $\lambda(\{q\}) = 0$ to the singletons and $\mathbb{Q}$ is countable.

5. (a) Prove *Chebychev inequality*: for a random variable X with $X(\omega) \geq 0$ P-almost surely,

$$\mathbb{P}(X > t) \leq \frac{E_{\mathbb{P}}\big(X\big)}{t} \quad \forall t > 0$$

Hint Note that

$$0 \leq t\, \mathbf{1}(X(\omega) > t) \leq X(\omega) \, .$$

(b) Prove *Chentsov inequality*

$$\mathbb{P}(X > t) \leq \inf_{\theta > 0} \Big\{ \exp(-\theta t) E_{\mathbb{P}}\big(\exp(\theta X)\big) \Big\}$$

Hint: for any $\theta > 0$, $X > t \Longleftrightarrow \exp(\theta X) > \exp(\theta t)$.

Solution Since the expectation is a positive operator, the inequality is preserved after taking expectation:

$$0 \le tP(X > t) \le E_P(X)$$

Let $Y = \exp(\theta X)$. Since the map $x \mapsto y = e^{\theta x}$ is strictly increasing when $\theta > 0$, seuraa $\forall \theta > 0$,

$$P(X > t) = P(Y > \exp(\theta X)) \le E_P(\exp(\theta X)) e^{-\theta t} \ .$$

(c) Consider a random variable $N(\omega)$ with Poisson(λ) distribution, where $\lambda > 0$ is the parameter and

$$\mathbb{P}_\lambda\big(N = k\big) = \exp(-\lambda)\frac{\lambda^k}{k!} \quad k \in \mathbb{N} = \{0, 1, 2, \dots\}$$

(d) Knowing that $E(\exp(\theta N)) = \exp\big(\lambda(e^\theta - 1)\big)$, (computed in the exercise sheet n.5) use Chentsov inequality to bound from above the probability $\mathbb{P}_\lambda\big(N > t\big)$, for $t > 0$.

$$P_\lambda\big(X > t\big) \le \inf_{\theta > 0}\Big\{\exp(-\theta t) E\big(\exp(\theta X)\big)\Big\} = \inf_{\theta > 0}\Big\{\exp\big(\lambda(e^\theta - 1) - \theta t\big)\Big\}$$

since at the minimum point t^*

$$\frac{\partial}{\partial \theta}\big(\lambda(e^\theta - 1) - \theta t\big) = \lambda e^\theta - t = 0$$

with

$$\frac{\partial^2}{\partial \theta^2}\big(\lambda(e^\theta - 1) - \theta t\big) = \lambda e^\theta > 0,$$

it follows that the function $\theta \mapsto \big(\lambda(e^\theta - 1) - \theta t\big)$ is convex and the minimum is achieved at $\theta^* = \log(t) - \log(\lambda)$, and we get the Chentsov upper bound as

$$P_\theta\big(X > t\big) \le \exp(t - \lambda)\left(\frac{\lambda}{t}\right)^t$$