STAT 712 hw 4

Transformations of a random variable, mgfs, quantile functions

Do problems 2.9, 2.12, 3.13(a), 3.14, 3.25, 3.26 from CB. In addition:

1. Let X be a random variable taking each value in its support $\mathcal{X} = \{x_1, \dots, x_n\}$ with probability $1/n$, and denote by $x_{(1)}, \dots, x_{(n)}$ the values $x_1, \dots, x_n$ when sorted such that $x_{(1)} < x_{(2)} < \dots < x_{(n)}$.

   (a) Give the quantile function $Q_X(u) = \inf\{x : F_X(x) \geq u\}$, $u \in (0, 1)$, where F_X is the cdf of X.

   Note that $F_X(x_{(i)}) = i/n$ for each $i = 1, \dots, n$. Now suppose $u \in [i/n, (i+1)/n)$ for some i. Then $\inf\{x : F_X(x) \geq u\} = x_{((i+1)/n)}$. Drawing a picture makes this clear. In order to define $Q_X(u)$ for all u, we can make use of the ceiling function $\lceil \cdot \rceil$ which returns the smallest integer greater than or equal to its argument (it "rounds up"). We have

   $$Q_X(u) = x_{(\lceil un \rceil)} \quad \text{for all } u \in (0, 1).$$

   (b) For some $h > 0$, let Y be a random variable with pdf

   $$f_Y(y) = \frac{1}{nh} \sum_{i=1}^{n} \phi\left(\frac{y - x_i}{h}\right),$$

   where ϕ is the pdf of the Normal$(0, 1)$ distribution.

   i. Show that f_Y is a valid pdf.

   We have

   $$\begin{aligned}
   \int_{-\infty}^{\infty} f_Y(y)dy &= \int_{-\infty}^{\infty} \frac{1}{nh} \sum_{i=1}^{n} \phi\left(\frac{y - x_i}{h}\right) dy \\
   &= \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \frac{1}{h} \phi\left(\frac{y - x_i}{h}\right) dy \\
   &= \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \phi(z_i)dz_i \quad (\text{ set } z_i = (y - x_i)/h\ ) \\
   &= 1.
   \end{aligned}$$

   ii. Show that the cdf F_Y of Y is given by

   $$F_Y(y) = \frac{1}{n} \sum_{i=1}^{n} \Phi\left(\frac{y - x_i}{h}\right) \quad \text{for all } y \in \mathbb{R},$$

   where Φ is the cdf of the Normal$(0, 1)$ distribution.

We have

$$
\begin{aligned}
F_Y(y) &= \int_{-\infty}^{y} \frac{1}{nh} \sum_{i=1}^{n} \phi\left(\frac{t - x_i}{h}\right) dt \\
&= \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{y} \frac{1}{h} \phi\left(\frac{t - x_i}{h}\right) dt \\
&= \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{(y-x_i)/h} \phi(z_i) dz_i \quad (\text{ set } z_i = (y - x_i)/h\ ) \\
&= \frac{1}{n} \sum_{i=1}^{n} \Phi\left(\frac{y - x_i}{h}\right).
\end{aligned}
$$

iii. Find $\mathbb{E}Y$.

We have

$$
\begin{aligned}
\mathbb{E}Y &= \int_{-\infty}^{\infty} y \cdot \frac{1}{nh} \sum_{i=1}^{n} \phi\left(\frac{y - x_i}{h}\right) dy \\
&= \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{\infty} y \cdot \frac{1}{h} \phi\left(\frac{y - x_i}{h}\right) dy \\
&= \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{\infty} (hz_i + x_i) \phi(z_i) dz_i \quad (\text{ set } z_i = (y - x_i)/h\ ) \\
&= \frac{1}{n} \sum_{i=1}^{n} x_i =: \bar{x}_n.
\end{aligned}
$$

iv. Find $\operatorname{Var} Y$.

We have

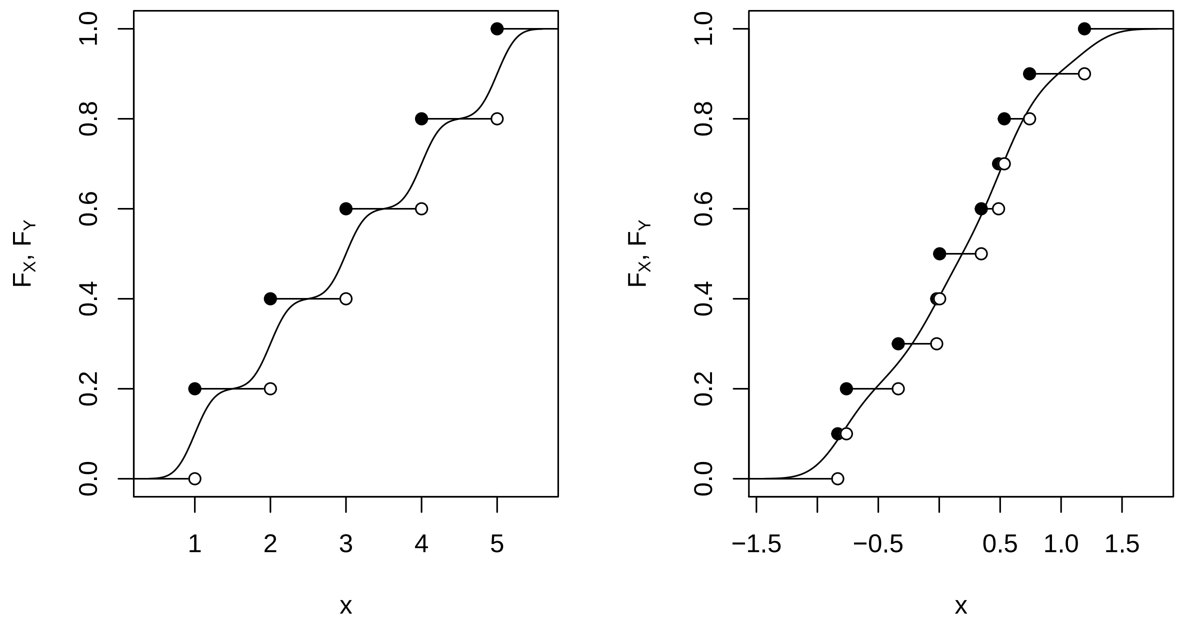
$$
\begin{aligned}
\mathbb{E}Y^2 &= \int_{-\infty}^{\infty} y^2 \cdot \frac{1}{nh} \sum_{i=1}^{n} \phi\left(\frac{y - x_i}{h}\right) dy \\
&= \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{\infty} (hz_i + x_i)^2 \phi(z_i) dz_i \quad (\text{ set } z_i = (y - x_i)/h\ ) \\
&= \frac{1}{n} \sum_{i=1}^{n} \left[h^2 \int_{-\infty}^{\infty} z_i^2 \phi(z_i) dz_i + 2hx_i \int_{-\infty}^{\infty} z_i \phi(z_i) dz_i + x_i^2 \int_{-\infty}^{\infty} \phi(z_i) dz_i \right] \\
&= h^2 + \frac{1}{n} \sum_{i=1}^{n} x_i^2.
\end{aligned}
$$

So

$$\operatorname{Var} Y = h^2 + \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2 + h^2.$$

(c) For $x_i = i$, $i = 1, 2, 3, 4, 5$, make a sketch of F_X with F_Y, for some small h, overlaid.

This can be done by hand, but here is a plot made in R that shows F_Y based on $x_i = i$, $i = 1, 2, 3, 4, 5$ in the left panel and F_Y based on $x_1, \dots, x_{10} \overset{\text{ind}}{\sim} \text{Normal}(0, 1)$ in the right panel.



We see that F_Y is a smoothed version of the empirical distribution function. The following code makes these plots:

```
library(latex2exp)
plot_edf_smooth_edf <- function(x,h){

  x <- sort(x)
  y.seq <- seq(min(x) - 3*h,max(x) + 3*h,length=200)
  plot(NA,
       xlim = range( y.seq ),
       ylim = c(0,1),
       ylab = TeX("$F_X$, $F_Y$"),
       xlab = TeX("$x$"))
```

```
  Fy <- numeric()
  for(i in 1:length(y.seq)){

    Fy[i] <- mean( pnorm((y.seq[i] - x)/h))

  }

  lines(Fy~y.seq)
  n <- length(x)
  x.tilde <- c(-10,x,10)
  for(i in 2:length(x.tilde)){

    lines(x = c(x.tilde[i-1],x.tilde[i]), y = c((i-2)/n,(i-2)/n))
    points(x = x.tilde[i], y = (i-2)/n,pch = 19,col = "white")
    points(x = x.tilde[i], y = (i-2)/n,pch = 1)
    points(x = x.tilde[i], y = (i-1)/n,pch = 19)

  }

}

par(mfrow = c(1,2))

x <- c(1,2,3,4,5)
plot_edf_smooth_edf(x,h=.2)

x <- rnorm(10)
plot_edf_smooth_edf(x,h=.2)
```

2. (Optional) Let F_X be a step function with jumps at the points $x_1, x_2, \dots$ and let $Q_X(u) = \inf\{x : F_X(x) \geq u\}$, $u \in (0,1)$. If $U \sim \text{Uniform}(0,1)$, show that the rv $X = Q_X(U)$ has cdf F_X.

Consider first, for some x_i,

$$P_X(X = x_i) = P_U(Q_X(U) = x_i) = P_U(F_X(x_{i-1}) < U \leq F_X(x_i)) = F_X(x_i) - F_X(x_{i-1}).$$

For the above, we can set $x_0 = -\infty$. We can see this by drawing pictures of a step function cdf and the corresponding quantile function. Now, for any $x \in \mathbb{R}$, we have

$$\begin{aligned} P(X \leq x) &= \sum_{\{i : x_i \leq x\}} P(X = x_i) \\ &= \sum_{\{i : x_i \leq x\}} [F_X(x_i) - F_X(x_{i-1})] \\ &= F_X(x_{i^*}) - F_X(x_0), \end{aligned}$$

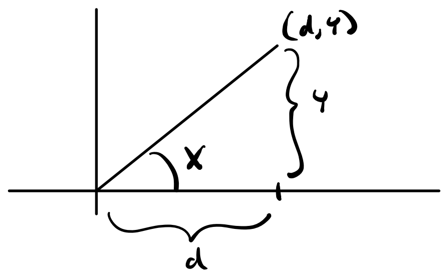
where $i^* = \max\{i : x_i \leq x\}$ and $F_X(x_0) = 0$. Since F_X is a step function, $F_X(x_{i^*}) = F_X(x)$, which shows that X has cdf F_X.

3. (Optional) 3.24 from CB is time-consuming, but is good for practicing transformations. I highly recommend working it out.

Problems 2.12, 3.13, 3.14, 3.25, 3.26 from CB.

$\boxed{2.12}$ Let $X \sim \text{Unif}(0, \pi/2)$, so $X \sim f_X(x) = \frac{2}{\pi} \mathbb{1}(0 < x < \pi/2)$

Let $Y = d \cdot \tan X$, from picture:



$\Rightarrow \quad \tan X = \frac{Y}{d}$

The support of Y is $(0, \infty)$, and we have

$$y = d \cdot \tan x = g(x) \iff x = \tan^{-1}\left(\frac{y}{d}\right) = g^{-1}(y)$$

$$\frac{d}{dy} g^{-1}(y) = \frac{1}{1 + y^2/d^2} \frac{1}{d},$$

The density of Y is given by

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| = \frac{2}{\pi} \frac{1}{d} \frac{1}{1 + y^2/d^2} \mathbb{1}(y > 0).$$

$\boxed{3.13}$

(a) Let X be a Poisson (λ) rv truncated such that $X > 0$. Then

(i) $$p_X(x) = \left(\frac{e^{-\lambda}\lambda^x}{x!}\right)\left(1 - \frac{e^{-\lambda}\lambda^0}{0!}\right)^{-1} \mathbb{1}(x \in \{1,2,\dots\})$$

$$= \left(\frac{e^{-\lambda}}{1-e^{-\lambda}}\right)\frac{\lambda^x}{x!}\,\mathbb{1}(x \in \{1,2,\dots\})$$

(ii) One way is to find the mgf:

$$M_X(t) = \sum_{x=1}^{\infty} e^{tx}\left(\frac{e^{-\lambda}}{1-e^{-\lambda}}\right)\frac{\lambda^x}{x!}$$

$$= \left(\frac{e^{-\lambda}}{1-e^{-\lambda}}\right)\sum_{x=1}^{\infty}\frac{(\lambda e^t)^x}{x!}$$

$$= \left(\frac{e^{-\lambda}}{1-e^{-\lambda}}\right)\left[\underbrace{\sum_{x=0}^{\infty}\frac{(\lambda e^t)^x}{x!}}_{= e^{\lambda e^t}} - \frac{(\lambda e^t)^0}{0!}\right]$$

$$= \frac{e^{\lambda(e^t-1)} - e^{-\lambda}}{1-e^{-\lambda}}.$$

Then

$$\mathbb{E}X = M_X^{(1)}(0)$$

$$= \frac{d}{dt}\,\frac{e^{\lambda(e^t-1)} - e^{-\lambda}}{1-e^{-\lambda}}\Bigg|_{t=0}$$

$$= \frac{1}{1-e^{-\lambda}} e^{\lambda(e^t-1)} \lambda e^t \Big|_{t=0}$$

$$= \frac{\lambda}{1-e^{-\lambda}} .$$

(iii) We have

$$\mathbb{E} X^2 = M_X^{(2)}(0)$$

$$= \frac{d}{dt}\left(\frac{1}{1-e^{-\lambda}} e^{\lambda(e^t-1)} \lambda e^t \right) \Big|_{t=0}$$

$$= \frac{1}{1-e^{-\lambda}} \left[e^{\lambda(e^t-1)} \lambda e^t \cdot \lambda e^t + e^{\lambda(e^t-1)} \lambda e^t \right] \Big|_{t=0}$$

$$= \frac{\lambda^2+\lambda}{1-e^{-\lambda}}$$

$$\operatorname{Var} X = \mathbb{E} X^2 - (\mathbb{E} X)^2 = \frac{\lambda^2+\lambda}{1-e^{-\lambda}} - \left(\frac{\lambda}{1-e^{-\lambda}}\right)^2 .$$

$\boxed{3.14}$ Let $X \sim p_X(x) = \dfrac{-(1-p)^x}{x \log p}$, $x = 1, 2, \ldots,$ $0 < p < 1.$

(a) Verify $\sum_{x=1}^{\infty} p_X(x) = 1.$

Expand $\log(1-t)$ around $t=0$: We have

$$\frac{d}{dt} \log(1-t) \Big|_{t=0} = \frac{-1}{1-t} \Big|_{t=0} = (-1)$$

$$\frac{d^2}{dt^2} \log(1-t) \Big|_{t=0} = \frac{-1(-1)(-1)}{(1-t)^2} \Big|_{t=0} = -\frac{1}{(1-t)^2} \Big|_{t=0} = -1$$

$$\frac{d^3}{dt^3} \log(1-t) \Big|_{t=0} = \frac{-2}{(1-t)^3} \Big|_{t=0} = -2$$

$$\frac{d^4}{dt^4} \log(1-t) \Big|_{t=0} = \frac{-3 \cdot 2}{(1-t)^4} \Big|_{t=0} = -3 \cdot 2$$

$$\vdots$$

$$\frac{d^k}{dt^k} \log(1-t) \Big|_{t=0} = -(k-1)! \quad , \quad \text{for } k \geq 2$$

So the Taylor expansion is

$$\log(1-t) = \sum_{k=0}^{\infty} \left(\frac{d^k}{dt^k} \log(1-t) \right) \Big|_{t=0} \frac{(t-0)^k}{k!}$$

$$= \log 2 + (-1)t - \sum_{k=2}^{\infty} \frac{(k-1)!\, t^k}{k!}$$

$$= -t - \frac{t^2}{2} - \frac{t^3}{3} - \cdots$$

$$= \sum_{k=1}^{\infty} -\frac{t^k}{k} .$$

Now we have $\log p = \log(1-(1-p)) = \sum_{k=1}^{\infty} -\frac{(1-p)^k}{k}$

$$\sum_{x=1}^{\infty} p_X(x) = \sum_{x=1}^{\infty} -\frac{(1-p)^x}{x \log p} = \frac{1}{\log p} \underbrace{\sum_{x=1}^{\infty} -\frac{(1-p)^x}{x}}_{\log p \ (\text{since } p = 1-(1-p))} = 1 .$$

(b) $$E X = \sum_{x=1}^{\infty} -\frac{x(1-p)^x}{x \log p}$$

$$= -\frac{1}{\log p} \sum_{x=1}^{\infty} (1-p)^x$$

see below

$$= -\frac{1}{\log p} \frac{1-p}{p}$$

$$= -\frac{(1-p)}{p \log p} .$$

Let $S_n = a + a^2 + \ldots + a^n$ and $a S_n = a^2 + a^3 + \ldots + a^{n+1}$.

Then $S_n - a S_n = a - a^{n+1} \Rightarrow S_n = \frac{a - a^{n+1}}{1-a}$.

Therefore

$$\sum_{x=1}^{\infty} a^x = \lim_{n\to\infty} S_n = \frac{a - \lim_{n\to\infty} a^{n+1}}{1-a}$$

Finding $\mathbb{E}X^2$ will be awkward unless we use funny tricks or the mgf. Let's use the mgf:

$$M_X(t) = \mathbb{E}\, e^{tX}$$

$$= \sum_{x=1}^{\infty} -\frac{(1-p)^x}{x \log p} \cdot e^{tx}$$

$$= \frac{1}{\log p} \sum_{x=1}^{\infty} \frac{-\left[(1-p)e^t\right]^x}{x}$$

$$= \frac{\log\left(1-(1-p)e^t\right)}{\log p}$$

$$\mathbb{E}X = M_X^{(1)}(0)$$

$$= \frac{d}{dt} \frac{\log\left(1-(1-p)e^t\right)}{\log p} \Bigg|_{t=0}$$

$$= \frac{1}{\log p} \frac{-(1-p)e^t}{1-(1-p)e^t} \Bigg|_{t=0}$$

$$= -\frac{(1-p)}{p \log p}$$

$$\mathbb{E}X^2 = M_X^{(2)}(0)$$

$$= \frac{d}{dt} \frac{1}{\log p} \frac{-(1-p)e^t}{1-(1-p)e^t} \Bigg|_{t=0}$$

$$= \frac{1}{\log p} \left[\frac{-(1-p)e^t}{[1-(1-p)e^t]^2} (-1)\left(-(1-p)e^t\right) + \frac{[-(1-p)e^t]}{1-(1-p)e^t} \right] \Bigg|_{t=0}$$

$$= \frac{1}{\log p} \left[\frac{-(1-p)^2}{p^2} - \frac{(1-p)}{p} \right]$$

$$= \frac{1}{\log p} \left[\frac{\overbrace{-(1-p)^2 - p(1-p)}^{-1+2p-p^2-p+p^2}}{p^2} \right]$$

$$= \frac{-(1-p)}{p^2 \log p}$$

g.

$$\operatorname{Var} X = \mathbb{E}X^2 - (\mathbb{E}X)^2$$

$$= \frac{-(1-p)}{p^2 \log p} - \left(\frac{-(1-p)}{p \log p} \right)^2$$

$$= \frac{-(1-p)}{p^2 \log p} \left[1 + \frac{(1-p)}{\log p} \right]$$

$\boxed{3.25}$ Let $T \sim f_T(t)$ with cdf F_T, continuous.

Hazard function is

$$h_T(t) = \lim_{\delta \downarrow 0} \frac{P(t \le T \le t+\delta \mid T \ge t)}{\delta}$$

$$= \lim_{\delta \downarrow 0} \frac{1}{\delta} \frac{P(\{t \le T \le t+\delta\} \cap \{T \ge t\})}{P(T \ge t)}$$

$$= \lim_{\delta \downarrow 0} \frac{1}{\delta} \frac{P(t \le T \le t+\delta)}{P(T \ge t)}$$

$$= \lim_{\delta \downarrow 0} \frac{\frac{1}{\delta}[F_T(t+\delta) - F_T(t)]}{1 - F_T(t)}$$

$$= \frac{f_T(t)}{1 - F_T(t)},$$

by the definition of a derivative.

Note

$$\frac{d}{dt}\left[-\log\left(1 - F_T(t)\right)\right] = \frac{f_T(t)}{1 - F_T(t)}.$$

by the chain rule.

$\boxed{3.26}$

(a) Let $T \sim \text{Exponential}(\beta)$, $\quad f_T(t) = \frac{1}{\beta} e^{-t/\beta} \mathbb{1}(t>0)$.

Then for $t>0$,

$$F_T(t) = \int_0^t \frac{1}{\beta} e^{-x/\beta} dx = -e^{-x/\beta} \Big|_0^t = 1 - e^{-t/\beta}.$$

So

$$h_T(t) = \frac{\frac{1}{\beta} e^{-t/\beta}}{1-\left(1-e^{-t/\beta}\right)} = \frac{1}{\beta}.$$

(b) Let $T \sim \text{Weibull}(\gamma, \beta)$, $\quad f_T(t) = \frac{\gamma}{\beta} t^{\gamma-1} e^{-t^\gamma/\beta} \mathbb{1}(x>0)$.

For $t>0$,

$$F_T(t) = \int_0^t \frac{\gamma}{\beta} x^{\gamma-1} e^{-\frac{x^\gamma}{\beta}} dx$$

$u = \frac{x^\gamma}{\beta}$ $\quad x = (\beta u)^{\frac{1}{\gamma}}$

$dx = \frac{1}{\gamma} (\beta u)^{\frac{1}{\gamma}-1} \beta \, du$

$$= \int_0^{t^\gamma/\beta} \frac{\gamma}{\beta} \left[(\beta u)^{\frac{1}{\gamma}}\right]^{\gamma-1} e^{-u} \frac{1}{\gamma} (\beta u)^{\frac{1}{\gamma}-1} \beta \, du$$

$$= \int_0^{t^\gamma/\beta} e^{-u} du$$

$$= -e^{-u} \Big|_0^{t^\gamma/\beta}$$

$$= 1 - e^{-t^\gamma/\beta}$$

So

$$h_T(t) = \frac{\frac{\gamma}{\beta} t^{\gamma-1} e^{-t^\gamma/\beta}}{1-\left(1-e^{-t^\gamma/\beta}\right)} = \frac{\gamma}{\beta} t^{\gamma-1}$$

(c) Let $T \sim \text{Logistic}(\mu, \beta)$, $F_T(t) = \dfrac{1}{1+e^{-(t-\mu)/\beta}}$.

The density is

$$f_T(t) = \frac{1}{\beta} \frac{e^{-(t-\mu)/\beta}}{\left[1+e^{-(t-\mu)/\beta}\right]^2},$$

So the hazard function is given by

$$h_T(t) = \frac{e^{-(t-\mu)\beta} \Big/ \left[1+e^{-(t-\mu)/\beta}\right]^2}{1-\dfrac{1}{1+e^{-(t-\mu)/\beta}}}$$

$$= \frac{e^{-(t-\mu)\beta} \Big/ \left[1+e^{-(t-\mu)/\beta}\right]^2}{\dfrac{e^{-(t-\mu)/\beta}}{1+e^{-(t-\mu)/\beta}}}$$

$$= \frac{1}{\beta} F_T(t).$$