Exercises on Distribution and Percentile Functions

H. Krieger, Mathematics 156, Harvey Mudd College

Fall, 2008

Throughout these exercises, unless otherwise specified, we let F be a distribution function and $R = F^{-1}$ be the corresponding percentile function.

1. If X has an exponential distribution with mean μ, find F_X and R_X explicitly.

2. Show that for arbitrary distribution function F, the percentile function R is always non-decreasing on $(0,1)$ and continuous from *below*.

3. Show that $p \leq F(R(p))$ for every $p \in (0,1)$ and that $R(F(x)) \leq x$ for every $x \in \mathbf{R}$.

4. Consequently, show that $R(F(R(p))) = R(p)$ for every $p \in (0,1)$ and that $F(R(F(x))) = F(x)$ for every $x \in \mathbf{R}$.

5. Show that if $X \sim F$ then, even if F is not continuous, $R(F(X)) \sim X$. Hint: $X \sim R(U)$ where $U \sim U(0,1)$.

6. For an explicit example of this, let

   $$P(X = -1) = 1/3 \text{ and } P(X = 2) = 2/3.$$

   Sketch the graphs of the corresponding distribution function F and percentile function R. Then verify that $X \sim R(U)$, where $U \sim U(0,1)$, and $X \sim R(F(X))$.

7. From above we know that if $p = F(x)$ for some $x \in \mathbf{R}$, *i.e.* p is in the range of F, then $F(R(p)) = p$. Show that this condition implies that p is a point of increase of R. Here, p is a point of increase of R if and only if for every $\varepsilon > 0$, $R(p+\varepsilon) - R(p-\varepsilon) > 0$. Consequently, show that every $p \in (0,1)$ is a point of increase of R if F is continuous.

8. Similarly, we know that if $x = R(p)$ for some $p \in (0,1)$, *i.e.* x is in the range of R, then $R(F(x)) = x$. Show that this condition implies that x is a point of increase of F.

9. Suppose that R is a non-decreasing function on $(0,1)$ which is continuous from below. We can extend R to $[0,1]$ by taking limits from within $(0,1)$ at the endpoints. To define an inverse for R, we proceed as follows. If $x < R(0)$ define $R^{-1}(x) = 0$. Otherwise, let $R^{-1}(x) = \max\{p : R(p) \leq x\}$. Show that R^{-1} is a distribution function. Also, show that if $R = F^{-1}$ for some distribution function F, then $R^{-1} = F$.

10. A four-parameter family of distributions can be defined for certain values of these parameters by the percentile function

$$R(p) = \lambda_1 + [p^{\lambda_3} - (1-p)^{\lambda_4}]/\lambda_2$$

for $p \in (0,1)$. Here, λ_1 is a location parameter, $\lambda_2 \neq 0$ is a scale parameter, and λ_3 and λ_4 are shape parameters (they determine skewness and kurtosis).

(a) Since R is differentiable, first show that

$$R'(p) = [\lambda_3 p^{\lambda_3 - 1} + \lambda_4 (1-p)^{\lambda_4 - 1}]/\lambda_2.$$

Then show that R is strictly increasing on $(0,1)$ if and only if $\lambda_3\lambda_4 \geq 0$, $\lambda_3 + \lambda_4 \neq 0$, and $\text{sign}(\lambda_2) = \text{sign}(\lambda_3 + \lambda_4)$.

(b) Assuming these conditions are satisfied, plot the density function (parametrically, using *e.g.* Maple, Matlab, Mathematica) for the distribution with $\lambda_1 = 0$, $\lambda_2 = 0.1975$, and $\lambda_3 = \lambda_4 = 0.1349$. Compare this plot with that of the standard normal density.

(c) Now suppose that $X \sim R(U)$ where $U \sim U(0,1)$. Then, assuming that these moments exist, show that

$$E([X - \lambda_1]^k) = \int_0^1 [R(p) - \lambda_1]^k dp = \lambda_2^{-k} A_k,$$

where

$$A_k = \sum_{j=0}^{k} \binom{k}{j} (-1)^j \beta(1 + (k-j)\lambda_3, 1 + j\lambda_4).$$

Here, the beta function is defined for $x > 0$ and $y > 0$ by

$$\beta(x,y) = \int_0^1 p^{x-1}(1-p)^{y-1} dp = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)},$$

where the gamma function (generalized factorial) is defined for $x > 0$ by

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt.$$

This function has the properties $\Gamma(1) = 1$ and $x\Gamma(x) = \Gamma(x+1)$ so that $\beta(x,y) = \beta(y,x)$ and $\beta(x,1) = 1/x$.

(d) Show that when k is a positive integer, A_k exists if $\min\{\lambda_3, \lambda_4\} > -1/k$.

(e) Evaluate A_k for $k = 0, 1, 2$. Also, show that $m_1 = E(X) = \lambda_2^{-1} A_1 + \lambda_1$ and $\sigma^2 = \mu_2 = E([X - m_1]^2) = \lambda_2^{-2}(A_2 - A_1^2)$.

(f) Show that $A_2 - A_1^2 \geq 0$ and thus we can solve to obtain $\lambda_2 = \lambda_2^*/\sigma$, where $\lambda_2^* = \text{sign}(\lambda_3 + \lambda_4)\sqrt{A_2 - A_1^2}$ is the value of λ_2 assuming $\sigma^2 = 1$.

(g) Similarly, if λ_1^* is the value of λ_1 assuming $m_1 = 0$ and $\sigma^2 = 1$, *i.e.* $\lambda_1^* = -A_1/\lambda_2^*$, show that $\lambda_1 = m_1 + \lambda_1^* \sigma$.