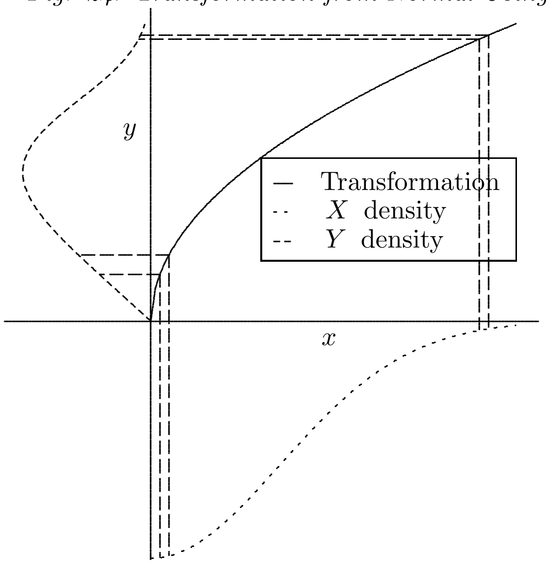
3. Derivative factor adjusts for local concentrating and diluting.
   a. See Fig. 24.

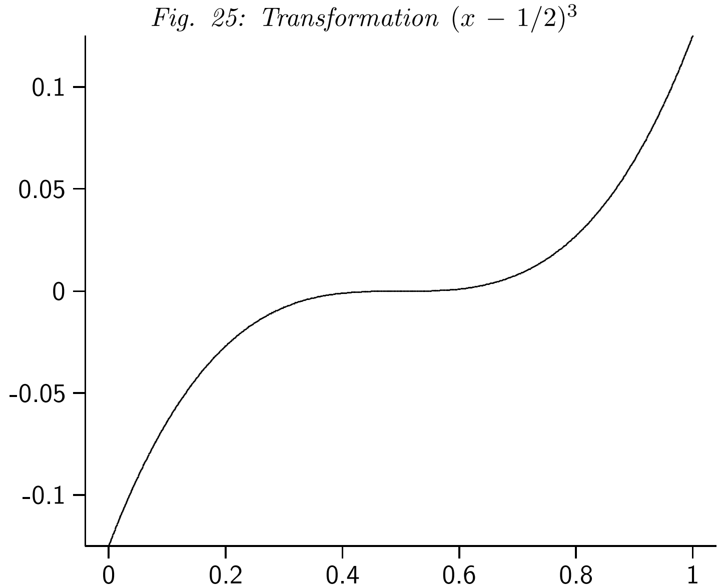
Fig. 24: Transformation from Normal Using Square Root



4. Simplest transformation example
   a. If $r(x) = cx$ for some constant c
      i. (as when the new measure is the old measure on a new scale)

   b. then the new probability density function is $f_Y(y) = f_X(y/c)/|c|$.
      i. This is as expected, since when the width of the range increases, the height of the function must decrease proportionally to compensate and leave the total integral one.
5. Argument requires that $dr^{-1}(y)/dy$ exists on $\mathcal{Y}$.
   a. $\frac{d}{dy}r^{-1}(y) = 1/r'(r^{-1}(y))$.
      i. By differentiating $r(r^{-1}(y)) = y$ gives $r'(r^{-1}(y))\frac{d}{dy}r^{-1}(y) = 1$.
   b. A transformation of a continuous variable can have a discrete distribution:
      i. X is uniform on $(0, 1)$ (ie., the probability density function is 1 throughout this region) and $r(x) = 0$.
      ii. Y is now discrete, taking only the value 0,
      iii. the arguments above break down because although the derivative of r exists, it is zero everywhere, and hence the derivative of r^{-1} exists nowhere.
6. Argument extends to some cases when r has a flat spot:
   a. Requires some more care.
   b. Example:
      i. X uniform on $(0, 1)$ and $r(x) = (x - .5)^3$. See Fig. 25.
      ii. Then $r'(x) = 3(x - .5)^2$
      iii. Then $r^{-1}(y) = y^{1/3} + .5$ and its derivative is $\frac{1}{3}y^{-2/3}$.
      iv. The probability density function over the domain $(-.5^{1/3}, .5^{1/3})$ is $1/(3(y^{1/3} + .5 - .5)^2) = \frac{1}{3}y^{-2/3}$.

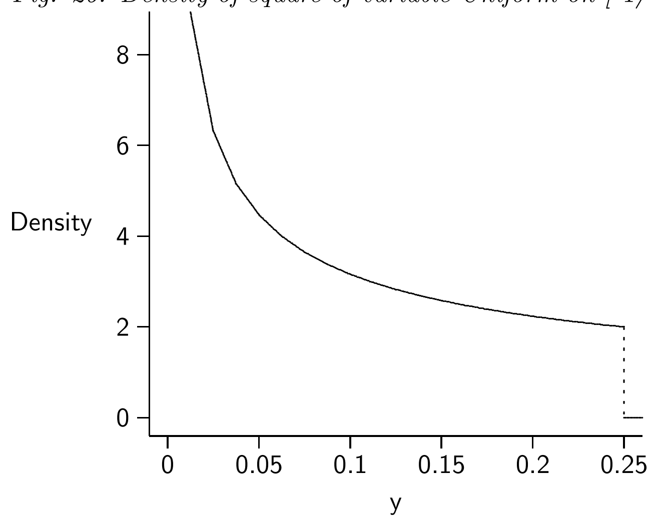
Fig. 25: Transformation $(x - 1/2)^3$



         - The probability density function is defined everywhere on the domain except at zero.
7. Case with transformation both increasing and decreasing:
   a. Domain $\mathcal{X}$ of X splits into disjoint subsets $\mathcal{X}_j$.
   b. r monotonic on each of $\mathcal{X}_j$.
      i. *monotonic* means either non-increasing, or non-decreasing.
   c. $f_Y(y) = \sum_{x \in r^{-1}(\{y\})} f_X(x)/r'(x)$.
   d. Example:

      i. X has probability density function $f_X(x)$ equalling 1 on $(-1/2, 1/2)$, 0 elsewhere
         - Transformation $Y = r(X)$ for $r(x) = x^2$
           ▷ non-increasing on $\mathcal{X}_1 = (-\frac{1}{2}, 0]$, and
           ▷ non-decreasing on $\mathcal{X}_2 = (0, \frac{1}{2})$.
         - $r^{-1}(y) = \sqrt{|y|}$
         - $\frac{d}{dy}r^{-1}(y) = \text{sgn}(y)\frac{1}{2}y^{-1/2}$.
         - $\text{sgn}(y) = \begin{cases} 1 & \text{if } y > 0 \\ -1 & \text{if } y < 0 \\ 0 & \text{if } y = 0 \end{cases}$.
         - $f_Y(y) = |-1/(2\sqrt{y})| + |1/(2\sqrt{y})| = y^{-1/2}$ for $0 < y \leq 1/4$, and 0 otherwise.
      ii. Note that formula fails at $y = 0$. See Fig. 26.
         - Remember from Riemann integration discussion that value of probability density function at one point doesn't matter.
      iii. The probability density function diverges to ∞ as $y \to 0$.
      iv. Integrals representing probabilities of sets like $(0, b]$ or $[0, b]$ are improper.
      v. Improper integrals are evaluated as limits of well-defined integrals.
         - $\text{P}(Y \leq 1/8) = \int_0^{1/8} y^{-1/2}\, dy$
         - Integral up to point where probability density function is infinite is taken as limit $\lim_{a \to 0, a>0} \int_a^{1/8} y^{-1/2}\, dy = \lim_{a \to 0, a>0} 2y^{1/2}\big|_a^{1/8} =$

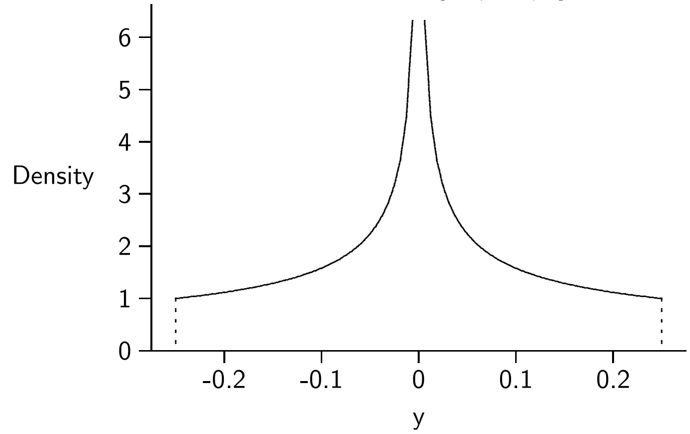
Fig. 26: Density of square of variable Uniform on [-1/2,1/2]



$\lim_{a\to 0, a>0} \sqrt{1/2} - 2\sqrt{a} = \sqrt{1/2}$.

8. probability density function may diverge to ∞ in middle.
   a. X has probability density function $f_X(x)$ equalling 1 on $(-1/2, 1/2)$, 0 elsewhere
   b. Transformation $Y = r(X)$ for $r(x) = \mathrm{sgn}(x)x^2$
      i. $r^{-1}(y) = \mathrm{sgn}(y)\sqrt{|y|}$
      ii. $\frac{d}{dy}r^{-1}(y) = \frac{1}{2}|y|^{-1/2}$.
      iii. $f_Y(y) = |1/(2\sqrt{|y|})|$ for $|y| \le 1/2$, and 0 otherwise. See Fig 27.
      iv. Formula still fails at $y = 0$.

Fig. 27: Density of signed square of variable uniform on $[-1/2, 1/2]$



      v. The probability density function still diverges to infinity as $y \to 0$.
         - Integrals representing probabilities of sets including 0 are improper.
   c. Can be addressed using ideas previously described.
      i. splitting at point where probability density function is problematic.
      ii. Treat the two improper bits as above.

WMS: 4.3

D. The expectation, mean, or average value.
1. Expectation Definition
   a. For continuous distńs is $\int_{\mathcal{X}} x f_X(x)\, dx$

   b. Don't define expectation if $\int_{\mathcal{X}} |x| f_X(x)\, dx = \infty$.
2. Expectation of transformation of a random variable defined as before
   a. Want $\mathsf{E}(r(X))$ for some random variable X taking values in $\mathcal{X}$.
   b. Transform to new variable $Y = r(X)$ taking values in $\mathcal{Y}$.
   c. Calculate its probability density function $f_Y(y)$
   d. Report $\mathsf{E}(Y) = \int_{\mathcal{Y}} y f_Y(y)\, dy$.
   e. Can calculate expectation of transformation without constructing new density: $\mathsf{E}(r(X)) = \int_{\mathcal{X}} r(x) f_X(x)\, dx$.
      i. As before, $f_Y(y) = \left|\frac{dx}{dy}\right| f_X(r^{-1}(y))$
      ii. $\int_{\mathcal{Y}} y f_X(r^{-1}(y)) \frac{dx}{dy}\, dy = \int_{\mathcal{X}} r(x) f_X(x)\, dx$
3. Definition of typical value
   a. Expectation
      i. Advantage: explicitly and uniquely defined.
      ii. Disadvantage: Sometimes isn't defined.
   b. Median
      i. Advantage: Always defined.
      ii. Disadvantage: Sometimes not unique.
4. Linearity
   a. Let $Y = aX + b$ for some constants a, b
   b. Then $\mathsf{E}(Y) = a\mathsf{E}(X) + b$.
      i. Use summation and constant multiple rules for integration:

$$\begin{aligned}\mathsf{E}(Y) &= \int_{\mathcal{X}} (ax+b) f_X(x)\, dx \\ &= a\int_{\mathcal{X}} x f_X(x)\, dx + b\int_{\mathcal{X}} f_X(x)\, dx = a\mathsf{E}(X) + b\end{aligned}$$

.

5. Other moments as before:
   a. The r*-th moment* is defined as $\mathsf{E}(X^r)$.
   b. The r*-th central moment* is defined as $\mathsf{E}((X - \mathsf{E}(X))^r)$.
6. Describing spread via *Variance*:
   a. $\mathsf{V}(X)$ is the second central moment: average squared distance from mean.
   b. Alternate formulation:
   $$\mathsf{V}(X) = \mathsf{E}\left(X^2\right) - \mathsf{E}(X)^2.$$
   c. *standard deviation*: typical distance from expectation: $\mathsf{SD}(X) = \sqrt{\mathsf{V}(X)}$
   d. Linearity:
   $$\begin{aligned}\mathsf{V}(aX+b) &= \mathsf{E}\left((aX + b - \mathsf{E}(aX+b))^2\right) \\ &= \mathsf{E}\left((aX + b - \mathsf{E}(aX) - b)^2\right) \\ &= \mathsf{E}\left(a^2(X - \mathsf{E}(X))^2\right) \\ &= a^2\mathsf{V}(X)\end{aligned}$$
      i. Hence $\mathsf{SD}(aX+b) = |a|\,\mathsf{SD}(X)$

WMS: 4.4

E. Particular Distributions
1. Uniform distribution
   a. In symbols, $X \sim \mathsf{Unif}(a,b)$.

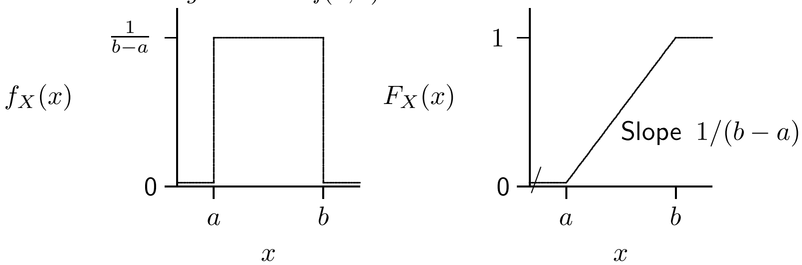
b. probability density function

$$f_X(x) = \begin{cases} 1/(b-a) & \text{if } x \in [a,b] \\ 0 & \text{otherwise.} \end{cases}$$

c. distribution function $F_X(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{x-a}{b-a} & \text{if } x \in [a,b] \\ 1 & \text{if } x > b. \end{cases}$

i. See Fig. 28.

Fig. 28: Unif(a, b) Distribution



Vertical scale on two panels is not the same.

d. Expectation $\mathsf{E}(X) = (a+b)/2$.

i. $\mathsf{E}(X) = \int_a^b x/(b-a)\, dx = (b^2/2 - a^2/2)/(b-a) = (a+b)/2$.

ii. We could have seen this through symmetry.

iii. Median is the same.

e. Variance: $\mathsf{V}(X) = (b-a)^2/12$.

i. $\mathsf{E}\left(X^2\right) = \int_a^b x^2/(b-a)\, dx = (b^3/3 - a^3/3)/(b-a) = (a^2+ab+b^2)/3$,

ii. $\mathsf{V}(X) = (a^2+ab+b^2)/3 - (a^2+2ab+b^2)/4 = (a^2-2ab+b^2)/12 = (b-a)^2/12$.

f. R gives probabilities via `punif`, but this is hardly necessary.

11

This page intentionally left blank.

This page intentionally left blank.