Chapter 3

Examples of Mass Functions and Densities

Inequalities
Outline

Jensen’s Inequality
  Convexity
  Examples

Markov’s Inequality
  Chebyshev’s Inequality

Chernoff Bound

Chernoff Bound
  Examples
Convexity

A set $C$ is **convex**, if for every pair of points $x_1, x_2 \in C$, the line segment $\alpha x_1 + (1 - \alpha) x_2, \alpha \in [0, 1]$, is a subset of $C$.

- A **real-valued function** $g$ is **convex** on $C$ if for every pair of points $x_1, x_2 \in C$,
  \[ g(\alpha x_1 + (1 - \alpha) x_2) \leq \alpha g(x_1) + (1 - \alpha)g(x_2) \]

- $g$ is **strictly convex** on $C$ if for $0 < \lambda < 1$,
  \[ g(\alpha x_1 + (1 - \alpha) x_2) < \alpha g(x_1) + (1 - \alpha)g(x_2) \]

- If $g$ is differentiable, then $g$ is **convex** if and only if the graph of $g$ is above the tangent line (hyperplane)
  \[ g(x) - g(x_0) \geq g'(x_0)(x - x_0) \quad g(x) - g(x_0) \geq \nabla g(x_0) \cdot (x - x_0) \text{ for all } x_0 \in C \]
Convexity

• If $g$ is differentiable, then $g$ is strictly convex if and only if the graph of $g$ is above the tangent line (hyperplane)

$$g(x) - g(x_0) > g'(x_0)(x - x_0) \quad g(x) - g(x_0) > \nabla g(x_0) \cdot (x - x_0) \quad \text{for all } x \neq x_0$$

Consequently, if $g'(x_0) = 0$ ($\nabla g(x_0) = 0$), then $g(x_0)$ is a global minimum.

• If $g$ is twice differentiable, then $g$ is convex if and only if

$$g''(x_0) \geq 0 \quad \text{Hessian matrix } Hg(x_0) \text{ is positive semidefinite for all } x_0 \in C.$$ 

$Hg$ is the the matrix of second order partial derivatives with $i, j$ entry

$$Hg(x)_{i,j} = \frac{\partial^2 g}{\partial x_i \partial x_j}(x).$$

with every vector $v$

$$v^T Hg(x) v \geq 0.$$
Jensen’s Inequality

Let $X = (X_1, \ldots, X_d)$ be a $d$-dimensional random vector with mean vector

$$\mu = (\mu_1, \ldots, \mu_d) = (EX_1, \ldots, EX_d).$$

If we look at the value $x_0 = \mu$ for the convex function $g$, then

$$g(x) - g(\mu) \geq \nabla g(\mu) \cdot (x - \mu).$$

Now replace $x$ with the random variable $X$ and take expectations.

$$E[g(X) - g(\mu)] \geq E[\nabla g(\mu) \cdot (X - \mu)] = \nabla g(\mu) \cdot E[X - \mu] = 0.$$ 

Consequently,

$$Eg(X) \geq g(\mu) = g(EX), \quad \text{Jensen’s inequality}$$
Examples

- Take $g(x) = x^2$, then
  $$EX^2 \geq (EX)^2,$$
  which follows from the fact that $\text{Var}(X) \geq 0$.
- Take $g(x) = x^{p/q}, p > q$ and $X = |Y|^q$, then
  $$E|Y|^p \geq (E|Y|^q)^{p/q}$$
  and
  $$\|Y\|_p = (E|Y|^p)^{1/p} \geq (E|Y|^q)^{1/q} = \|Y\|_q$$
  The $L_p$-norm is an increasing function of $p$.
- Take $g(x) = 1/x$ and $X$ positive, then
  $$E[1/X] \geq 1/EX, \quad EX \geq 1/E[1/X],$$
  The mean is greater than the harmonic mean.
Examples

\[ g(x_1, x_2) = -x_1^\lambda x_2^{1-\lambda}, \quad 0 < \lambda < 1, \quad x_1, x_2 \geq 0, \]
then \( g \) is convex. Thus, by Jensen's inequality,
\[ E[X_1^\lambda X_2^{1-\lambda}] \leq (EX_1)^\lambda (EX_2)^{1-\lambda} \]

For \( \lambda = 1/p, \quad 1 - \lambda = 1/q \), (Thus, \( 1/p + 1/q = 1 \).) \( X_1 = |Y_1|^p, \quad X_2 = |Y_2|^q \),
\[ E[|Y_1 Y_2|] \leq ||Y_1||_p ||Y_2||_q \]
Hölder’s inequality.
Markov’s Inequality

For a positive function $g$ define

$$m_B = \inf\{g(t); t \in B\}$$

Then

$$Eg(X) \geq E[g(X)I_B(X)] \geq E[m_B I_B(X)] = m_B P\{X \in B\}.$$  

Markov’s inequality occurs by taking $g$ to be function increasing on the support of $X$ and $B = (x, \infty)$, then $m_B = g(x)$, and

$$Eg(X) \geq g(x)P\{X > x\} \quad \text{and} \quad P\{X > x\} \leq \frac{Eg(X)}{g(x)}.$$
Markov’s Inequality

For the case of a non-negative random variable $X$ and $g(x) = x$. The area of the rectangle $xP\{X > x\}$ is less than $EX$, the area above the graph of the cumulative distribution function and below the horizontal line at 1.

Figure: A geometric proof of Markov’s inequality $xP\{X > x\} \leq EX$
Chebyshev’s Inequality

The case $X = |Y - \mu_Y|$ and $g(x) = x^2$ is called Chebyshev’s inequality. In this case, we have

$$P\{|Y - \mu_Y| > y\} \leq \frac{E(Y - \mu_Y)^2}{y^2} = \frac{\text{Var}(Y)}{y^2}.$$

For a standard normal random variable $Z$, this yields

$$P\{|Z| \geq z\} \leq \frac{1}{z^2}.$$

The exact value is much less.

```r
> z<-2:5; data.frame(z,1/z^2,2*pnorm(-z))
   z    X1.z.2 X2...pnorm..z.
1 2 0.2500000 4.550026e-02
2 3 0.1111111 2.699796e-03
3 4 0.0625000 6.334248e-05
4 5 0.0400000 5.733031e-07
```
Chernoff Bound

If we choose \( g(x) = \exp(tx), t > 0 \), then for a random variable \( X \) possessing a moment generating function \( M_X(t) \), the Markov inequality becomes

\[
P\{X > x\} = P\{\exp tX > \exp tx\} \leq \frac{M_X(t)}{\exp(tx)},
\]

\[
\ln P\{X > x\} \leq \ln M_X(t) - tx = K_X(t) - tx.
\]

where \( K_X(t) \) is the cumulant generating function.
Next, we minimize this inequality over all possible choices of $t$.

$$\ln P\{X > x\} \leq -K^*(x)$$

or

$$P\{X > x\} \leq \exp(-K^*(x))$$

where $\inf_{t>0}\{K_X(t) - tx\} = -\sup_{t>0}\{tx - K_X(t)\} = -K^*_X(x)$.  

To find supremum, we first show $K_X(t) - tx$ has a unique maximum by verifying that $K_X(t)$ is strictly convex.
Chernoff Bound

Thus, choose $\alpha \in (0, 1)$.

\[ K_X(\alpha t_1 + (1 - \alpha)t_2) = \ln M_X(\alpha t_1 + (1 - \alpha)t_2) \]
\[ = \ln E[\exp(\alpha t_1 + (1 - \alpha)t_2)X] \]
\[ = \ln E[\exp((\alpha t_1 X) \exp((1 - \alpha)t_2 X))] \]
\[ < \ln E[\exp(t_1 X)]^\alpha E[\exp(t_2 X)]^{1 - \alpha} \]
\[ = \alpha \ln E[\exp(t_1 X)] + (1 - \alpha) \ln E[\exp(t_2 X)] \]
\[ = \alpha K_X(t_1) + (1 - \alpha)K_X(t_2) \]

The inequality follows from Hölder’s inequality. Thus, $K_X(t)$ is strictly convex.
Chernoff Bound

\( K^*_X(x) \) is called the (convex) conjugate function, or the Legendre transform for \( K_X \). For this application, \( K^*_X \) is also called rate function.

Because \( tx - K_X(t) \) is strictly concave down, the maximum of \( tx - K_X(t) \) is unique. Take a derivative with respect to \( t \) and set the expression equal to 0 to obtain

\[
K'_X(t) = x
\]

Let \( t^*(x) \) denote the solution to this equation. Then,

\[
K^*_X(x) = t^*(x)x - K_X(t^*(x)).
\]
Chernoff Bound

- The cumulant generating function $K_X$ is a convex function of $t$.
- Add a plot of a line with slope $x$.
- At the maximum difference, the slope of the tangent line to $K_X$ is $x$.
- Call this value $t^*(x)$.
- The Legendre transform $K^*_X(x)$ is the difference in the values of the cumulant generating function and the line.

$$K^*_X(x) = t^*(x)x - K_X(t^*(x))$$
Chernoff Bound

Example. For the standard normal, the cumulant generating function, $K_Z(t) = t^2/2$. Thus,

$$K'_Z(t) = t, \quad t^*(x) = x, \quad K^*(x) = t^*(x)x - K_X(t^*(x)) = x^2 - \frac{x^2}{2} = \frac{x^2}{2}.$$ 

Thus, for $z > 0$, $P\{Z > z\} \leq \exp\left(-\frac{z^2}{2}\right)$.

Returning to our estimates:

```r
g z <- (z, 2*pnorm(-z), exp(-z^2/2))
z x2...pnorm..z. exp..z.2.2.
1 2 4.550026e-02 1.353353e-01
2 3 2.699796e-03 1.110900e-02
3 4 6.334248e-05 3.354626e-04
4 5 5.733031e-07 3.726653e-06
5 6 1.973175e-09 1.522998e-08
```
Chernoff Bound

Example. For a Poisson random variable, the moment generating function, $M_X(t) = \rho_X(e^t) = \exp(\lambda(e^t - 1))$ and the cumulant generating function $K_X(t) = \lambda(e^t - 1)$.

$$K_X'(t) = \lambda e^t, \quad t^*(x) = \ln \frac{x}{\lambda}, \quad K^*(x) = t^*(x)x - K_X(t^*(x)) = x \ln \frac{x}{\lambda} - x + \lambda.$$  

Thus, for $x > 0$, $P\{X > x\} \leq \exp -K^*(x) = \left(\frac{\lambda}{x}\right)^x e^{x-\lambda}$.

Take $\lambda = 5$ and look at 3 to 5 standard deviations above the mean.

```r
> x<-seq(15,25,2);data.frame(x,1-ppois(x,5),(5/x)^x*exp(x-5))
   x X1...ppois.x..5. X.5.x..x...exp.x...5.
1 15 6.900824e-05 1.535062e-03
2 17 5.416338e-06 1.501040e-04
3 19 3.452136e-07 1.159403e-05
4 21 1.820653e-08 7.252314e-07
5 23 8.072533e-10 3.748612e-08
6 25 3.049971e-11 1.627944e-09
```