556: MATHEMATICAL STATISTICS I

SOME INEQUALITIES

Expectation Inequalities

JENSEN'S INEQUALITY

Jensen's Inequality gives a lower bound on expectations of convex functions. Recall that a function $g(x)$ is **convex** if, for $0 < \lambda < 1$,

$$g(\lambda x + (1-\lambda)y) \leq \lambda g(x) + (1-\lambda)g(y)$$

for all x and y. Alternatively, if the derivatives are well defined, function $g(x)$ is **convex** if

$$\frac{d^2}{dt^2}\left\{g(t)\right\}_{t=x} = g^{(2)}(x) \geq 0.$$

Conversely, $g(x)$ is **concave** if $-g(x)$ is convex.

Theorem (JENSEN'S INEQUALITY)

Suppose that X is a random variable with expectation μ, and function g is convex and finite. Then

$$\mathbb{E}_X\left[g(X)\right] \geq g(\mathbb{E}_X\left[X\right])$$

with equality if and only if, for every line $a + bx$ that is a tangent to g at μ

$$\mathrm{P}_X[g(X) = a + bX] = 1.$$

that is, $g(x)$ is linear.

Proof Let $l(x) = a + bx$ be the equation of the tangent at $x = \mu$. Then, for each x, $g(x) \geq a + bx$ as in the figure. Thus

$$\mathbb{E}_X[g(X)] \geq \mathbb{E}_X[a + bX] = a + b\mathbb{E}_X[X] = l(\mu) = g(\mu) = g(\mathbb{E}_X[X])$$

as required. Also, **if** $g(x)$ is linear, then equality follows by properties of expectations. Suppose that

$$\mathbb{E}_X\left[g(X)\right] = g(\mathbb{E}_X\left[X\right]) = g(\mu)$$

but $g(x)$ is convex, but not linear. Let $l(x) = a + bx$ be the tangent to g at μ. Then by convexity

$$g(x) - l(x) > 0 \qquad \therefore \qquad \int (g(x) - l(x))\, \mathrm{d}F_X(x) = \int g(x)\, \mathrm{d}F_X(x) - \int l(x)\, \mathrm{d}F_X(x) > 0$$

and hence

$$\mathbb{E}_X[g(X)] > \mathbb{E}_X[l(X)].$$

But $l(x)$ is linear, so $\mathbb{E}_X[l(X)] = a + b\mathbb{E}_X[X] = g(\mu)$, yielding the contradiction

$$\mathbb{E}_X[g(X)] > g(\mathbb{E}_X[X]).$$

and the result follows.

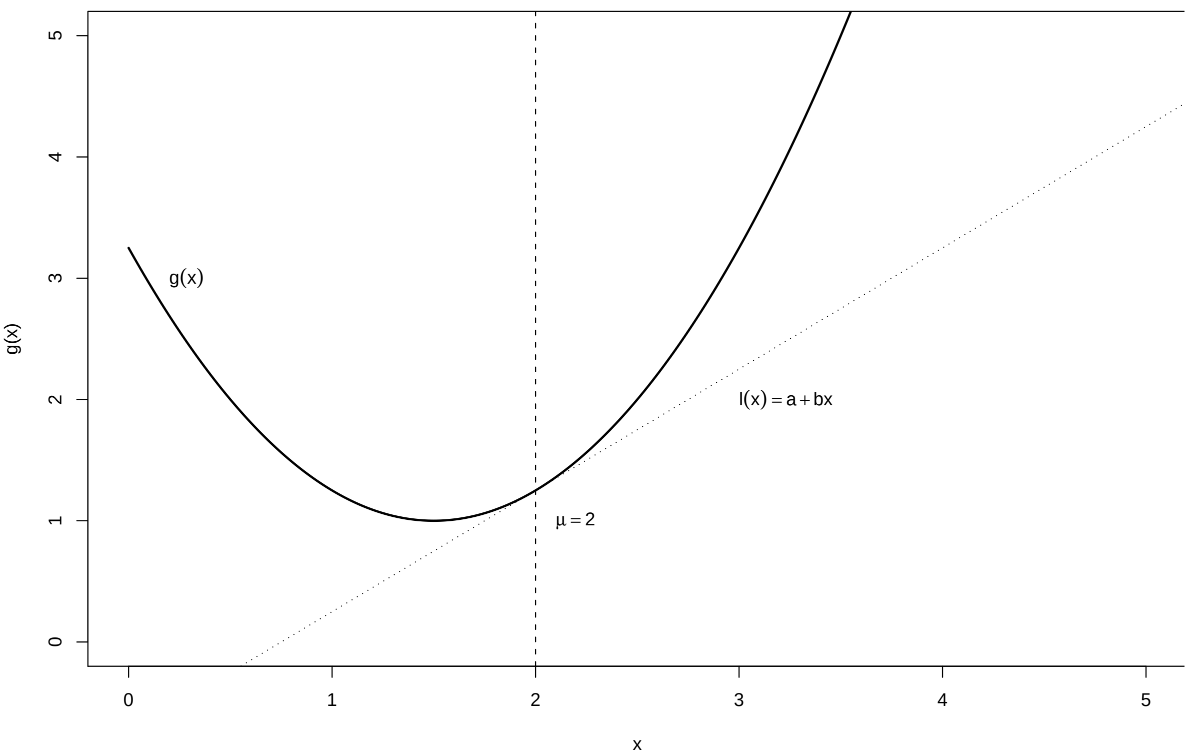

Figure 1: The function $g(x)$ and its tangent at $x = \mu$.

- If $g(x)$ is **concave**, then
$$\mathbb{E}_X\left[g(X)\right] \leq g(\mathbb{E}_X\left[X\right])$$
- $g(x) = x^2$ is **convex**, thus
$$\mathbb{E}_X\left[X^2\right] \geq \{\mathbb{E}_X\left[X\right]\}^2$$
- $g(x) = \log x$ is **concave**, thus
$$\mathbb{E}_X\left[\log X\right] \leq \log\{\mathbb{E}_X\left[X\right]\}$$

Alternative approach to Jensen's Inequality:
We may use the general definition of convexity to prove the result by using the fact that the distribution F_X can be viewed as a limiting function derived from a sequence of discrete cdfs. We have that $g(x)$ is convex if, for $n \geq 2$ and constants $\lambda_j, j = 1, \ldots, n$, with $0 < \lambda_j < 1$, and $\lambda_1 + \cdots + \lambda_n = 1$

$$g\left(\sum_{j=1}^{n} \lambda_j x_j\right) \leq \sum_{j=1}^{n} \lambda_j g\left(x_j\right)$$

for all vectors $(x_1, \ldots, x_n)$; this follows by induction using the original definition. We may regard this statement as stating

$$g\left(\mathbb{E}_n[X]\right) \leq \mathbb{E}_n[g(X)] \tag{1}$$

where

$$\mathbb{E}_n[X] = \int x \; \mathrm{d}F_n(x) \qquad \mathbb{E}_n[g(X)] = \int g(x) \; \mathrm{d}F_n(x)$$

where F_n is the cdf of the discrete distribution on $\{x_1, \ldots, x_n\}$ with associated probability masses $\{\lambda_1, \ldots, \lambda_n\}$, that is,

$$F_n(x) = \sum_{j=1}^{n} \lambda_j \mathbb{I}_{[x_j,\infty)}(x).$$

Now, for any F_X, we can find infinite sequences $\{(x_j, \lambda_j), j = 1, 2, \ldots\}$ such that for all x

$$\lim_{n \longrightarrow \infty} F_n(x) = F_X(x)$$

– this is stated pointwise here, but convergence functionwise also holds. Also, as g is convex, it is also continuous. Therefore we may pass limits through the integrals and note that

$$\lim_{n \longrightarrow \infty} \mathbb{E}_n[X] = \mathbb{E}_X[X] \qquad \lim_{n \longrightarrow \infty} \mathbb{E}_n[g(X)] = \mathbb{E}_X[g(X)]$$

which yields Jensen's inequality by substitution into (1).

CAUCHY-SCHWARZ INEQUALITY

Theorem
For random variable X and functions $g_1()$ and $g_2()$, we have that

$$\{\mathbb{E}_X[g_1(X)g_2(X)]\}^2 \leq \mathbb{E}_X[\{g_1(X)\}^2]\mathbb{E}_X[\{g_2(X)\}^2] \tag{2}$$

with equality if and only if either $\mathbb{E}_X[\{g_1(X)\}^2] = 0$ or $\mathbb{E}_X[\{g_2(X)\}^2] = 0$, or

$$\mathrm{P}_X[g_1(X) = cg_2(X)] = 1$$

for some $c \neq 0$.

Proof Let $X_1 = g_1(X)$ and $X_2 = g_2(X)$, and let

$$Y_1 = aX_1 + bX_2 \qquad Y_2 = aX_1 - bX_2$$

and as $\mathbb{E}_{Y_1}[Y_1^2], \mathbb{E}_{Y_2}[Y_2^2] \geq 0$, we have that

$$a^2\mathbb{E}_X[X_1^2] + b^2\mathbb{E}_X[X_2^2] + 2ab\mathbb{E}_X[X_1X_2] \geq 0$$

$$a^2\mathbb{E}_X[X_1^2] + b^2\mathbb{E}_X[X_2^2] - 2ab\mathbb{E}_X[X_1X_2] \geq 0$$

Set $a^2 = \mathbb{E}_X[X_2^2]$ and $b^2 = \mathbb{E}_X[X_1^2]$. If either a or b is zero, the inequality clearly holds. We may thus consider $\mathbb{E}_X[X_1^2], \mathbb{E}_X[X_2^2] > 0$: we have

$$2\mathbb{E}_X[X_1^2]\mathbb{E}_X[X_2^2] + 2\{\mathbb{E}_X[X_1^2]\mathbb{E}_X[X_2^2]\}^{1/2}\mathbb{E}_X[X_1X_2] \geq 0$$

$$2\mathbb{E}_X[X_1^2]\mathbb{E}_X[X_2^2] - 2\{\mathbb{E}_X[X_1^2]\mathbb{E}_X[X_2^2]\}^{1/2}\mathbb{E}_X[X_1X_2] \geq 0$$

Rearranging, we obtain that

$$-\{\mathbb{E}_X[X_1^2]\mathbb{E}_X[X_2^2]\}^{1/2} \leq \mathbb{E}_X[X_1X_2] \leq \{\mathbb{E}_X[X_1^2]\mathbb{E}_X[X_2^2]\}^{1/2}$$

that is $\{\mathbb{E}_X[X_1X_2]\}^2 \leq \mathbb{E}_X[X_1^2]\mathbb{E}_X[X_2^2]$ or, in the original form

$$\{\mathbb{E}_X[g_1(X)g_2(X)]\}^2 \leq \mathbb{E}_X[\{g_1(X)\}^2]\mathbb{E}_X[\{g_2(X)\}^2].$$

We examine the case of equality:

$$\{\mathbb{E}_X[g_1(X)g_2(X)]\}^2 = \mathbb{E}_X[\{g_1(X)\}^2]\mathbb{E}_X[\{g_2(X)\}^2] \tag{3}$$

If $\mathbb{E}_X[\{g_j(X)\}^2] = 0$ for $j = 1$ or 2, then $g_j(X)$ is constant with probability one, say $\mathrm{P}_X[g_j(X) = c] = 1$. Clearly the left-hand side of (2) is non-negative, so we must have equality as the right-hand side is zero. So suppose $\mathbb{E}_X[\{g_j(X)\}^2] > 0$ for $j = 1, 2$, but $g_1(X) = cg_2(X)$ with probability one for some $c \neq 0$. In this case we replace $g_1(X)$ in the left- and right- hand sides of (2) to conclude that

$$\{\mathbb{E}_X[cg_2(X)^2]\}^2 = \mathbb{E}_X[\{cg_2(X)\}^2]\mathbb{E}_X[\{g_2(X)\}^2] = c^2\mathbb{E}_X[\{g_2(X)\}^2]$$

and equality follows.

For the converse, assume that (3) holds. If both sides equate to zero, then we must have at least one term on the right-hand side equal to zero, so $\mathbb{E}_X[\{g_j(X)\}^2] = 0$ for $j = 1$ or 2. If both sides equate to a positive constant then both $\mathbb{E}_X[\{g_j(X)\}^2] > 0$. By assumption, we may write

$$\mathbb{E}_X[\{g_1(X)\}^2] = \frac{\{\mathbb{E}_X[g_1(X)g_2(X)]\}^2}{\mathbb{E}_X[\{g_2(X)\}^2]}$$

say. Let $Z = g_1(X) - cg_2(X)$. For a contradiction, assume that Z is not zero with probability 1: we have

$$\mathbb{E}[Z^2] = \mathbb{E}[\{g_1(X)\}^2] + c^2\mathbb{E}[\{g_2(X)\}^2] - 2c\mathbb{E}[g_1(X)g_2(X)]$$

which is strictly positive. However the right hand side can be written,

$$\mathbb{E}[\{g_1(X)\}^2] + \left(c\{\mathbb{E}[\{g_2(X)\}^2]\}^{1/2} - \frac{\mathbb{E}[g_1(X)g_2(X)]}{\{\mathbb{E}[\{g_2(X)\}^2]\}^{1/2}}\right)^2 - \left(\frac{\mathbb{E}[g_1(X)g_2(X)]}{\{\mathbb{E}[\{g_2(X)\}^2]\}^{1/2}}\right)^2$$

Now if we set

$$c = \frac{\mathbb{E}[g_1(X)g_2(X)]}{\mathbb{E}[\{g_2(X)\}^2]}$$

the second term is zero, so we must then have

$$\mathbb{E}[\{g_1(X)\}^2] - \frac{\{\mathbb{E}[g_1(X)g_2(X)]\}^2}{\mathbb{E}[\{g_2(X)\}^2]} > 0$$

but this contradicts assumption (3). Hence Z must be zero with probability 1, that is

$$g_1(X) = cg_2(X)$$

with probability 1.

HÖLDER'S INEQUALITY

Lemma Let $a, b > 0$ and $p, q > 1$ satisfy

$$p^{-1} + q^{-1} = 1. \tag{4}$$

Then

$$p^{-1}\,a^p + q^{-1}\,b^q \geq ab$$

with equality if and only if $a^p = b^q$.

Proof Fix $b > 0$. Let

$$g(a;b) = p^{-1}\,a^p + q^{-1}\,b^q - ab.$$

We require that $g(a; b) \geq 0$ for all a. Differentiating wrt a for fixed b yields $g^{(1)}(a; b) = a^{p-1} - b$, so that $g(a; b)$ is minimized (the second derivative is strictly positive at all a) when $a^{p-1} = b$, and at this value of a, the function takes the value

$$p^{-1}\, a^p + q^{-1}\, (a^{p-1})^q - a(a^{p-1}) = p^{-1}\, a^p + q^{-1}\, a^p - a^p = 0$$

as, by equation (4), $1/p + 1/q = 1 \Longrightarrow (p-1)q = p$. As the second derivative is strictly positive at all a, the minimum is attained at the **unique** value of a where $a^{p-1} = b$, where, raising both sides to power q yields $a^p = b^q$.

Theorem (HÖLDER'S INEQUALITY)
Suppose that X and Y are two random variables, and $p, q > 1$ satisfy (4). Then

$$|\mathbb{E}_{X,Y}[XY]| \leq \mathbb{E}_{X,Y}[|XY|] \leq \{\mathbb{E}_X[|X|^p]\}^{1/p} \{\mathbb{E}_{f_Y}[|Y|^q]\}^{1/q}$$

Proof (Absolutely continuous case: discrete case similar) For the first inequality,

$$\mathbb{E}_{X,Y}[|XY|] = \iint |xy| f_{X,Y}(x,y)\; dx\; dy \geq \iint xy f_{X,Y}(x,y)\; dx\; dy = \mathbb{E}_{X,Y}[XY]$$

and

$$\mathbb{E}_{X,Y}[XY] = \iint xy f_{X,Y}(x,y)\; dx\; dy \geq \iint -|xy| f_{X,Y}(x,y)\; dx\; dy = -\mathbb{E}_{X,Y}[|XY|]$$

so

$$-\mathbb{E}_{X,Y}[|XY|] \leq \mathbb{E}_{X,Y}[XY] \leq \mathbb{E}_{X,Y}[|XY|] \qquad \therefore \qquad |\mathbb{E}_{X,Y}[XY]| \leq \mathbb{E}_{X,Y}[|XY|].$$

For the second inequality, set

$$a = \frac{|X|}{\{\mathbb{E}_X[|X|^p]\}^{1/p}} \qquad b = \frac{|Y|}{\{\mathbb{E}_{f_Y}[|Y|^q]\}^{1/q}}.$$

Then from the previous lemma

$$p^{-1}\, \frac{|X|^p}{\mathbb{E}_X[|X|^p]} + q^{-1}\, \frac{|Y|^q}{\mathbb{E}_{f_Y}[|Y|^q]} \geq \frac{|XY|}{\{\mathbb{E}_X[|X|^p]\}^{1/p} \{\mathbb{E}_{f_Y}[|Y|^q]\}^{1/q}}$$

and taking expectations yields, on the left hand side,

$$p^{-1}\, \frac{\mathbb{E}_X[|X|^p]}{\mathbb{E}_X[|X|^p]} + q^{-1}\, \frac{\mathbb{E}_{f_Y}[|Y|^q]}{\mathbb{E}_{f_Y}[|Y|^q]} = p^{-1} + q^{-1} = 1$$

and on the right hand side

$$\frac{\mathbb{E}_{X,Y}[|XY|]}{\{\mathbb{E}_X[|X|^p]\}^{1/p} \{\mathbb{E}_{f_Y}[|Y|^q]\}^{1/q}}$$

and the result follows.

Note: here we have equality if and only if

$$\mathrm{P}_{X,Y}[|X|^p = c|Y|^q] = 1$$

for some non zero constant c.

Theorem (CAUCHY-SCHWARZ INEQUALITY REVISITED)
Suppose that X and Y are two random variables.

$$|\mathbb{E}_{X,Y}[XY]| \leq \mathbb{E}_{X,Y}[|XY|] \leq \left\{\mathbb{E}_X[|X|^2]\right\}^{1/2}\left\{\mathbb{E}_{f_Y}[|Y|^2]\right\}^{1/2}$$

Proof Set $p = q = 2$ in the Hölder Inequality.

Corollaries:

(a) Let μ_X and μ_Y denote the expectations of X and Y respectively. Then, by the Cauchy-Schwarz inequality

$$|\mathbb{E}_{X,Y}[(X-\mu_X)(Y-\mu_Y)]| \leq \left\{\mathbb{E}_X[(X-\mu_X)^2]\right\}^{1/2}\left\{\mathbb{E}_{f_Y}[(Y-\mu_Y)^2]\right\}^{1/2}$$

so that

$$\mathbb{E}_{X,Y}[(X-\mu_X)(Y-\mu_Y)] \leq \mathbb{E}_X[(X-\mu_X)^2]\mathbb{E}_{f_Y}[(Y-\mu_Y)^2]$$

and hence

$$\{\text{Cov}_{X,Y}[X,Y]\}^2 \leq \text{Var}_X[X]\,\text{Var}_{f_Y}[Y].$$

(b) **Lyapunov's Inequality**: Define $Y = 1$ with probability one. Then, for $1 < p < \infty$

$$\mathbb{E}_X[|X|] \leq \{\mathbb{E}_X[|X|^p]\}^{1/p}.$$

Let $1 < r < p$. Then

$$\mathbb{E}_X[|X|^r] \leq \{\mathbb{E}_X[|X|^{pr}]\}^{1/p}$$

and letting $s = pr > r$ yields

$$\mathbb{E}_X[|X|^r] \leq \{\mathbb{E}_X[|X|^s]\}^{r/s}$$

so that

$$\{\mathbb{E}_X[|X|^r]\}^{1/r} \leq \{\mathbb{E}_X[|X|^s]\}^{1/s}$$

for $1 < r < s < \infty$.

Theorem (MINKOWSKI'S INEQUALITY)
Suppose that X and Y are two random variables, and $1 \leq p < \infty$. Then

$$\{\mathbb{E}_{X,Y}[|X+Y|^p]\}^{1/p} \leq \{\mathbb{E}_X[|X|^p]\}^{1/p} + \{\mathbb{E}_{f_Y}[|Y|^p]\}^{1/p}$$

Proof Write

$$\begin{aligned}\mathbb{E}_{X,Y}[|X+Y|^p] &= \mathbb{E}_{X,Y}[|X+Y||X+Y|^{p-1}] \\ &\leq \mathbb{E}_{X,Y}[|X||X+Y|^{p-1}] + \mathbb{E}_{X,Y}[|Y||X+Y|^{p-1}]\end{aligned}$$

by the triangle inequality $|x+y| \leq |x| + |y|$. Using Hölder's Inequality on the terms on the right hand side, for q selected to satisfy $1/p + 1/q = 1$,

$$\mathbb{E}_{X,Y}[|X+Y|^p] \leq \{\mathbb{E}_X[|X|^p]\}^{1/p}\left\{\mathbb{E}_{X,Y}[|X+Y|^{q(p-1)}]\right\}^{1/q} + \{\mathbb{E}_{f_Y}[|Y|^p]\}^{1/p}\left\{\mathbb{E}_{X,Y}[|X+Y|^{q(p-1)}]\right\}^{1/q}$$

and dividing through by $\left\{\mathbb{E}_{X,Y}[|X+Y|^{q(p-1)}]\right\}^{1/q}$ yields

$$\frac{\mathbb{E}_{X,Y}[|X+Y|^p]}{\left\{\mathbb{E}_{X,Y}[|X+Y|^{q(p-1)}]\right\}^{1/q}} \leq \{\mathbb{E}_X[|X|^p]\}^{1/p} + \{\mathbb{E}_{f_Y}[|Y|^p]\}^{1/p}$$

and the result follows as $q(p-1) = p$, and $1 - 1/q = 1/p$.

Concentration and Tail Probability Inequalities

Lemma (CHEBYCHEV'S LEMMA) If X is a random variable, then for non-negative function h, and $c > 0$,

$$\mathrm{P}_X\left[h(X) \geq c\right] \leq \frac{\mathbb{E}_X\left[h(X)\right]}{c}$$

Proof (continuous case) : Suppose that X has density function f_X which is positive for $x \in \mathbb{X}$. Let $\mathcal{A} = \{x \in \mathbb{X} : h(x) \geq c\} \subseteq X$. Then, as $h(x) \geq c$ on $\mathcal{A}$,

$$\begin{aligned}
\mathbb{E}_X\left[h(X)\right] = \int h(x) f_X(x)\, dx &= \int_{\mathcal{A}} h(x) f_X(x)\, dx + \int_{\mathcal{A}'} h(x) f_X(x)\, dx \\
&\geq \int_{\mathcal{A}} h(x) f_X(x)\, dx \\
&\geq \int_{\mathcal{A}} c f_X(x)\, dx = c\,\mathrm{P}_X\left[X \in \mathcal{A}\right] = c\,\mathrm{P}_X\left[h(X) \geq c\right]
\end{aligned}$$

and the result follows.

- **SPECIAL CASE I - THE MARKOV INEQUALITY**
  If $h(x) = |x|^r$ for $r > 0$, so

  $$\mathrm{P}_X\left[|X|^r \geq c\right] \leq \frac{\mathbb{E}_X\left[|X|^r\right]}{c}.$$

  Alternately stated (by Casella and Berger) as follows: If $P[Y \geq 0] = 1$ and $P[Y = 0] < 1$, then for any $r > 0$

  $$\mathrm{P}_Y[Y \geq r] \leq \frac{\mathbb{E}_X\left[Y\right]}{r}$$

  with equality if and only if

  $$\mathrm{P}_Y[Y = r] = p = 1 - \mathrm{P}_Y[Y = 0]$$

  for some $0 < p \leq 1$.

- **SPECIAL CASE II - THE CHEBYCHEV INEQUALITY**
  Suppose that X is a random variable with expectation μ and variance σ^2. Then $h(x) = (x - \mu)^2$ and $c = k^2\sigma^2$, for $k > 0$,

  $$\mathrm{P}_X\left[(X - \mu)^2 \geq k^2\sigma^2\right] \leq 1/k^2$$

  or equivalently

  $$\mathrm{P}_X\left[|X - \mu| \geq k\sigma\right] \leq 1/k^2.$$

  Setting $\epsilon = k\sigma$ gives

  $$\mathrm{P}_X\left[|X - \mu| \geq \epsilon\right] \leq \sigma^2/\epsilon^2$$

  or equivalently

  $$\mathrm{P}_X\left[|X - \mu| < \epsilon\right] \geq 1 - \sigma^2/\epsilon^2.$$

Theorem (TAIL BOUNDS FOR THE NORMAL DENSITY)
If $Z \sim \mathcal{N}(0,1)$, then for $t > 0$

$$\sqrt{\frac{2}{\pi}} \frac{t}{1+t^2}\, e^{-t^2/2} \le \mathrm{P}_Z[|Z| \ge t] \le \sqrt{\frac{2}{\pi}} \frac{1}{t} e^{-t^2/2}$$

Proof By symmetry, $\mathrm{P}_Z[|Z| \ge t] = 2\, \mathrm{P}_Z[Z \ge t]$, so

$$\mathrm{P}_Z[Z \ge t] \;=\; \left(\frac{1}{2\pi}\right)^{1/2} \int_t^\infty e^{-x^2/2}\, dx \le \left(\frac{1}{2\pi}\right)^{1/2} \int_t^\infty \frac{x}{t} e^{-x^2/2}\, dx = \left(\frac{1}{2\pi}\right)^{1/2} \frac{e^{-t^2/2}}{t}.$$

Similarly, for $t > 0$,

$$\int_t^\infty e^{-x^2/2}\, dx \equiv \int_t^\infty \frac{x}{x}\, e^{-x^2/2}\, dx = \left[-\frac{1}{x} e^{-x^2/2}\right]_t^\infty - \int_t^\infty \frac{1}{x^2}\, e^{-x^2/2}\, dx \ge \frac{1}{t} e^{-t^2/2} - \frac{1}{t^2} \int_t^\infty e^{-x^2/2}\, dx$$

after writing $1 = x/x$, then integrating by parts, and then noting that, on (t, ∞), $x > t \iff 1/x^2 < 1/t^2$, and that the integrand is non-negative. Therefore, combining terms

$$\left(1 + \frac{1}{t^2}\right) \int_t^\infty e^{-x^2/2}\, dx \ge \frac{1}{t}\, e^{-t^2/2}$$

and cross-multiplying by the positive term $t^2/(1+t^2)$ yields

$$\int_t^\infty e^{-x^2/2}\, dx \ge \frac{t}{1+t^2}\, e^{-t^2/2} \qquad \therefore \qquad \mathrm{P}_Z[|Z| > t] \ge \sqrt{\frac{2}{\pi}} \frac{t}{1+t^2}\, e^{-t^2/2}.$$

To see the quality of the approximation, the table below shows the values of the bounding values for t ranging from 1 to 5. Clearly the bounds improve as t gets larger.

t	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
Lower	2.420e-01	1.196e-01	4.319e-02	1.209e-02	2.659e-03	4.610e-04	6.298e-05	6.770e-06	5.718e-07
True	3.173e-01	1.336e-01	4.550e-02	1.242e-02	2.700e-03	4.653e-04	6.334e-05	6.795e-06	5.733e-07
Upper	4.839e-01	1.727e-01	5.399e-02	1.402e-02	2.955e-03	4.987e-04	6.692e-05	7.104e-06	5.947e-07