Math 425
Intro to Probability
Lecture 37

Kenneth Harris
kaharri@umich.edu

Department of Mathematics
University of Michigan

April 18, 2009

Example – continued

☞ Consider a Bernoulli trials process with IID indicator variables $X_1, X_2, \ldots$ denoting whether the trial was a success or failure. Suppose the probability of success is p. So,

$$E[X_i] = p \qquad Var(X_i) = p(1-p).$$

☞ Let $A_n = \frac{X_1 + \ldots + X_n}{n}$ be the sample average over n trials.

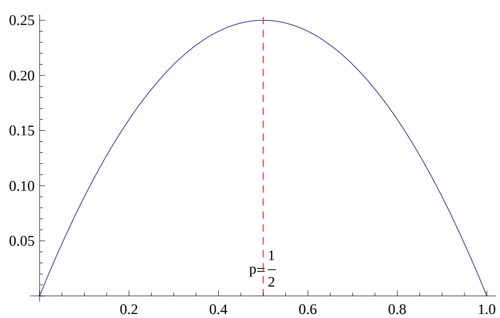
$$E\left[A_n\right] = p \qquad Var\left(A_n\right) = \frac{p(1-p)}{n}.$$

☞ By Chebyshev's inequality, for any $\varepsilon > 0$

$$\mathbf{P}\left\{\left|A_n - p\right| \geq \varepsilon\right\} \leq \frac{p(1-p)}{n\varepsilon^2}.$$

Example – continued

Example – continued. The variance $\sigma^2 = p(1-p)$ has a maximum value of $\frac{1}{4}$ achieved at $p = \frac{1}{2}$:



Plugging back into Chebyshev gives us a bound on the deviation of the sample average from the mean:

$$\mathbf{P}\left\{\left|A_n - p\right| \geq \varepsilon\right\} \leq \frac{p(1-p)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}.$$

Example

Example

We have two coins: one is fair and the other produces heads with probability $3/4$. One coin is picked at random. How many tosses suffice for us to be 95 percent sure of which coin we had?

☞ To make this problem more concrete: if the proportion of heads is less than 0.625, then we will guess the coin was fair; otherwise, if the proportion of heads is greater than 0.625 we will guess the biased coin.

How many tosses will suffice for 95 percent certainty that the generated sample average will not deviate by more than $\epsilon = 0.125$ from its mean?

Example – continued

☞ By Chebyshev's inequality:

$$\mathbf{P}\left\{\left|A_n - p\right| \geq \varepsilon\right\} \;\leq\; \frac{1}{4n\varepsilon^2}.$$

☞ We want to n large enough so that we have only 5% error:

$$\frac{1}{4n\varepsilon^2} \;\leq\; 0.05$$

equivalently,

$$n \;\geq\; \frac{1}{4(0.05)\varepsilon^2} = \frac{5}{\varepsilon^2}$$

☞ We now have a bound on the number of trials needed without needing to know the mean or the variance.

Example – continued

☞ For $\varepsilon = 0.125$ choose n so that

$$n \geq \frac{5}{(0.125)^2} = 320$$

☞ By tossing the coin $n \geq 320$ times we can be 95% certain the sample average is within 0.125 of the true bias p of the coin to heads:

$$\mathbf{P}\left\{\left|A_n - p\right| \geq 0.125\right\} \;\leq\; 0.05$$

☞ Toss the coin 320 times and count heads.

- If fewer than 200 heads appear guess the fair coin.
- If more than 200 heads appear guess the biases coin.
- If exactly 200 heads appears, then laugh at your (bad?) luck.

You can be 95 percent certain you chose the right coin.

Degree of Certainty vs. Number of Trials

☞ To achieve certainty p that we are within $\varepsilon = 0.125$ of the mean requires n trials, where

$$n \;\geq\; \frac{1}{4(1-p)(0.125)^2}$$

Degree of Certainty	Number of Trials
50%	32
75%	64
90%	160
95%	320
99%	1600
99.9%	16,000

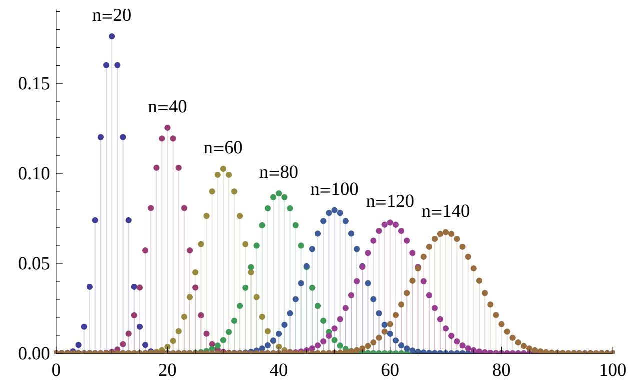
Sums of Random Variables

☞ Let $X_1, X_2, \ldots$ be IID (independent and identically distributed) random variables with a common mean μ and variance σ^2.

Let $S_n = X_1 + X_2 + \ldots + X_n$, so the statistics for S_n are

$$E[S_n] = n\mu \qquad Var(S_n) = n\sigma^2 \qquad StDev(S_n) = \sqrt{n}\sigma$$

So S_n is tending to shift to the right and flatten-out as $n \to \infty$.



Standardization

☞ We standardize the sums S_n to guarantee they have the same mean and variance.

Definition

Let $X_1, X_2, \ldots$ be IID random variables with a common mean μ and variance σ^2. Let $S_n = X_1 + X_2 + \ldots + X_n$.

The standardization of S_n is the random variable

$$S_n^* = \frac{S_n - n\mu}{\sqrt{n}\sigma}$$

Proposition

The statistics of the standardization S_n^* of S_n are

$$E[S_n^*] = 0 \qquad Var(S_n^*) = 1 \qquad \text{for any } n.$$

Proof of Proposition

☞ Let $X_1, X_2, \ldots$ be IID random variables with a common mean μ and variance σ^2. Let $S_n = X_1 + X_2 + \ldots + X_n$, so that the statistics of S_n are

$$E[S_n] = n\mu \qquad Var(S_n) = n\sigma^2$$

☞ The standardization S_n^* is defined as

$$S_n^* = \frac{S_n - n\mu}{\sqrt{n}\sigma}$$

so its statistics are

$$E[S_n^*] = \frac{E[S_n] - n\mu}{\sqrt{n}\sigma} = 0$$

$$Var(S_n^*) = Var\left(\frac{S_n - n\mu}{\sqrt{n}\sigma}\right) = \frac{Var(S_n)}{n\sigma^2} = 1.$$

Central Limit Theorem

Theorem (Central Limit Theorem, CLT)

Let $X_1, X_2, \ldots$ be a sequence of IID random variables having mean μ and variance σ^2. Let $S_n = X_1 + X_2 + \ldots + X_n$ and S_n^* be its standardization

$$S_n^* = \frac{S_n - n\mu}{\sqrt{n}\sigma}$$

☞ *Then the distribution of the random variable S_n^* tends to the standard normal distribution as $n \to \infty$.*

That is, for any $-\infty < a < \infty$

$$\mathbf{P}\{S_n^* \leq a\} \longrightarrow \Phi(a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2}\, dx \qquad \text{as } n \to \infty.$$

Convergence of the Central Limit Theorem

☞ The CLT only states that for each a

$$\mathbf{P}\{S_n^* \leq a\} \longrightarrow \Phi(a),$$

where $\Phi(a)$ is the standard normal distribution.

CLT leaves open a couple important issues

1. How large n must be for $\Phi(a)$ to be close to $\mathbf{P}\{S_n^* \leq a\}$,
2. Is there a single n which works for all a; or, will n vary with each a?

☞ It would be very BAD if it turned-out n was usually very large, or even if the tightness of the approximation for a choice of n depended on a.

Convergence of the Central Limit Theorem

☞ Fortunately, this does not happen in MOST circumstances. The following result states that the convergence in CLT is on the order of $\frac{1}{\sqrt{n}}$ independently of a.

☞The Berry-Esseen Theorem states

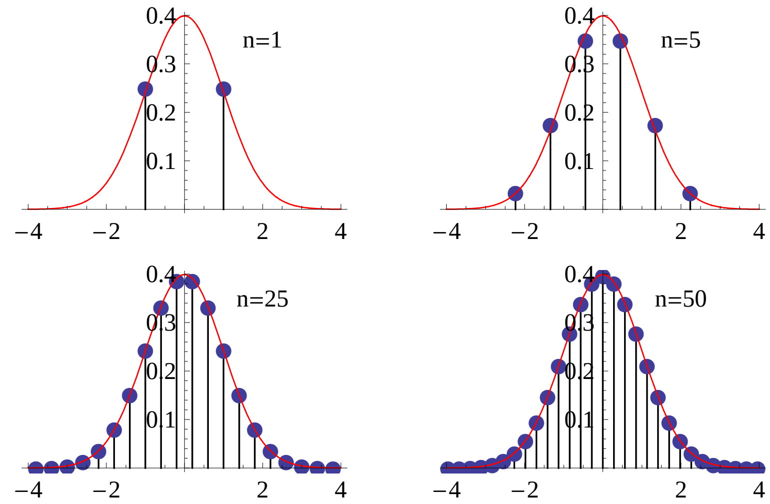
If $X_1, X_2, \ldots$ are IID random variables with finite mean μ, variance σ^2, and third moment $E[|X_i|^3]$, then there is a constant C (which does not depend on a or n) such that for any real a and integer n

$$\left|\mathbf{P}\left\{S_n^* \le a\right\} \; - \; \Phi(a)\right| \; \le \; \frac{C}{\sqrt{n}}.$$

☞ The rule of thumb is that the central limit theorem provides a good approximation when $n \geq 30$.

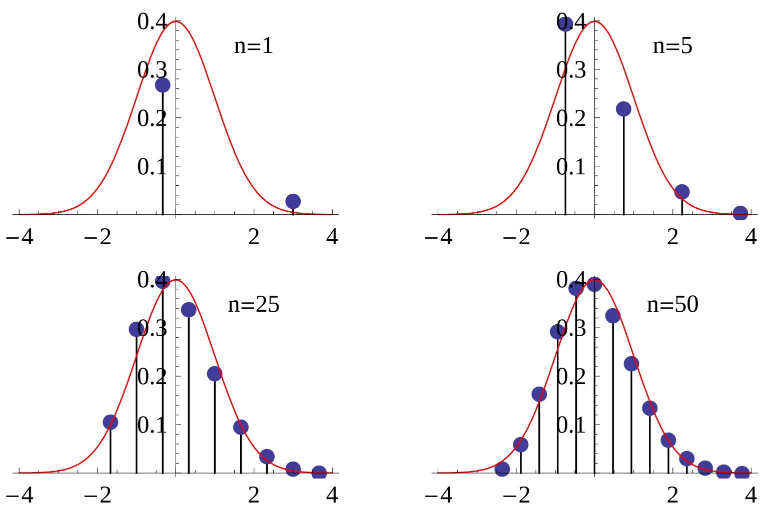
Binomial Distribution – $p = 0.5$

☞ Central Limit Theorem for Bernoulli random variable with $p = 0.5$.



Binomial Distribution – $p = 0.1$

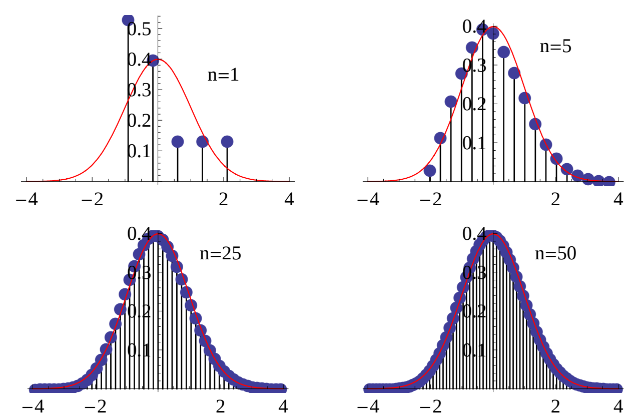
☞ Central Limit Theorem for Bernoulli random variable with $p = 0.1$.



Arbitrary Discrete Distribution

☞ Central Limit Theorem for an arbitrary discrete distribution

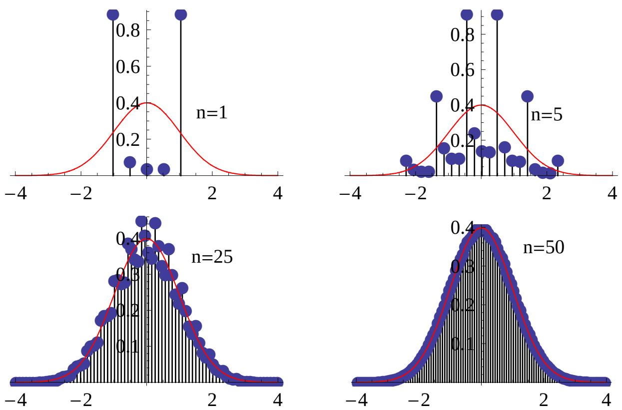
$$X = \begin{pmatrix} x & 0 & 1 & 2 & 3 & 4 \\ p & 0.4 & 0.3 & 0.1 & 0.1 & 0.1 \end{pmatrix}$$



Arbitrary Discrete Distribution

☞ Central Limit Theorem for an arbitrary discrete distribution

$$X = \begin{pmatrix} x & 0 & 1 & 2 & 3 & 4 \\ p & 0.46 & 0.04 & 0.02 & 0.02 & 0.46 \end{pmatrix}$$



Standardized Continuous Density

☞ Let $X_1, X_2, \ldots$ be IID continuous random variables with common mean μ and variance σ. We can compute the density $f_{S_n}(x)$ for the sums $S_n = X_1 + X_2 + \ldots + X_n$ using the convolution.

☞ To compute the density $f_{S_n^*}(x)$ for standardized sum S_n^*

$$S_n^* = \frac{S_n - n\mu}{\sqrt{n}\sigma}$$

use Theorem 5.7.1 (Ross, page 243):

$$\begin{aligned} F_{S_n^*}(x) &= F_{S_n}(x\sqrt{n}\sigma + n\mu) \\ f_{S_n^*}(x) &= \sqrt{n}\sigma \cdot f_{S_n}(x\sqrt{n}\sigma + n\mu). \end{aligned}$$

Example

Example. Let $X_1, X_2, \ldots$ be IID exponential random variables with common mean parameter λ. So,

$$E[X_i] = \frac{1}{\lambda} \qquad Var(X_i) = \frac{1}{\lambda^2}.$$

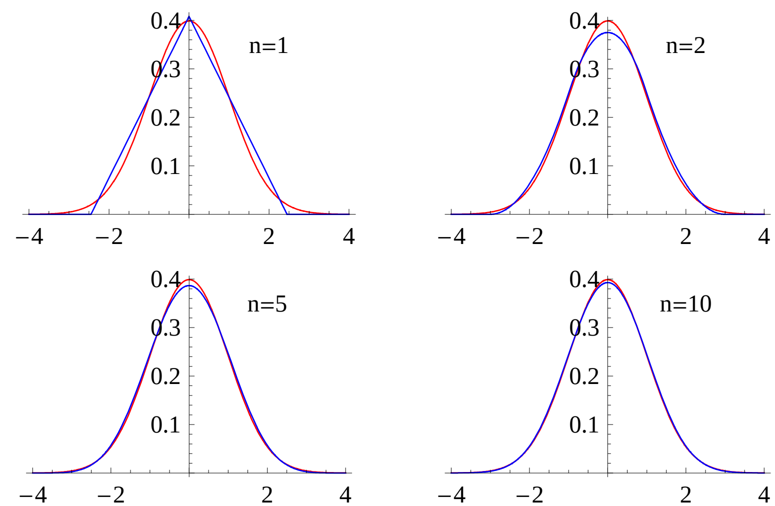
☞ Recall that the sum of n exponential random variables with parameter λ is a gamma distributed random variable with parameters (n, λ) (Section 6.3 of Ross).

☞ The density for S_n^* in this case is

$$\begin{aligned} f_{S_n}(x) &= \frac{\lambda e^{-\lambda x}(\lambda x)^{n-1}}{(n-1)!} \\ f_{S_n^*}(x) &= \frac{\sqrt{n}}{\lambda} \cdot f_{S_n}\left(\frac{\sqrt{n}x + n}{\lambda}\right). \end{aligned}$$

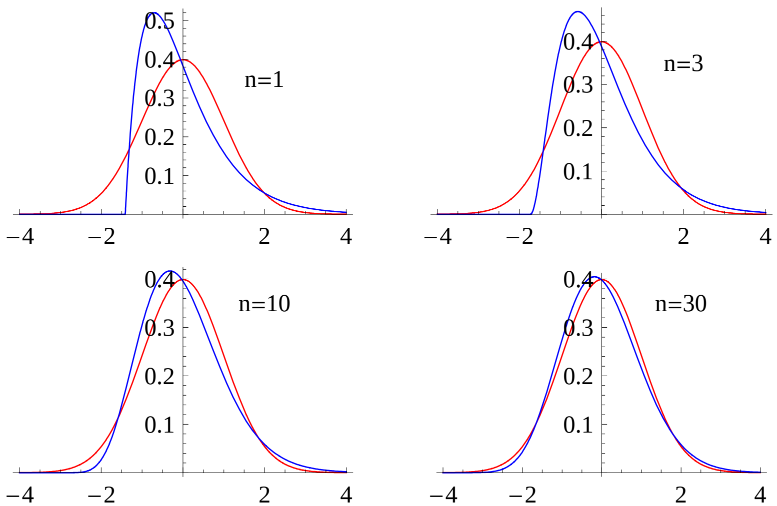
Uniform Distribution

☞ Central Limit Theorem for the uniform distribution on $[0, 1]$ plotted against the standard normal distribution.



Exponential Distribution

☞ Central Limit Theorem for the exponential distribution with parameter $\lambda = 1$ plotted against the standard normal distribution.



Example

Example

Fifty numbers are rounded-off to the nearest integer and the summed. Suppose that the individual round-off errors are uniformly distributed over $(-0.5, 0.5)$. What is the probability that the round-off error exceeds the exact sum by more than 3?

Solution. Let X_i $(i = 1, \dots, 50)$ be the round-off error on the i number, and $X = X_1 + X_2 + \dots + X_{50}$ the total round-off error.
The X_i are IID and uniformly distributed, so

$$E[X_i] = 0 \qquad Var(X_i) = \frac{1}{12}$$
$$E[X] = 0 \qquad Var(X) = \frac{50}{12}$$

☞ Apply CLT to approximate the probability $\mathbf{P}\{|X| > 3\}$

Example – continued

☞ First standardize E, then apply the normal approximation

$$\begin{aligned}
\mathbf{P}\{|E| > 3\} &= \mathbf{P}\{|X^*| > \frac{3-0}{\sqrt{50/12}}\} \\
&= \mathbf{P}\{|X^*| > \frac{6\sqrt{3}}{5\sqrt{2}}\} \\
&= \mathbf{P}\{X^* > \frac{6\sqrt{3}}{5\sqrt{2}}\} + \mathbf{P}\{X^* < -\frac{6\sqrt{3}}{5\sqrt{2}}\} \\
&\approx 2 - 2 \cdot \Phi(\frac{6\sqrt{3}}{5\sqrt{2}}) \\
&\approx 2 - 2(0.9292) \\
&= 0.1416.
\end{aligned}$$

Example

Example

A student's grade is the average of 30 assignments, where each assignment is recorded as an integer out of 100 possible points. Suppose that the instructor makes an error in grading of $\pm k$ with probability $|\varepsilon/k|$, where $|k| \le 5$ and $\varepsilon = \frac{1}{20}$. The distribution of errors is

$$\begin{pmatrix} k & 0 & \pm 1 & \pm 2 & \pm 3 & \pm 4 & \pm 5 \\ p & \frac{463}{600} & \frac{1}{20} & \frac{1}{40} & \frac{1}{60} & \frac{1}{80} & \frac{1}{100} \end{pmatrix}$$

The final grade is obtained by averaging the 30 assignment grades.
What is the probability that the difference between the "correct average grade" and the recorded average grade differs by less than 0.05 for a given student?

Example – continued

☞ Let X_i ($i = 1, 2, \ldots, 30$) be the error between the actual score on the ith assignment and the recorded score. We will assume the errors are independent. Let $X = X_1 + X_2 + \ldots + X_{30}$, the sum of the errors.

The statistics are computed from the distribution

$$E[X_i] = 0 \qquad Var(X_i) = 1.5$$
$$E[X] = 0 \qquad Var(X) = (30)1.5 = 45.$$

☞ Apply CLT to approximate the probability

$$\mathbf{P}\{-0.05 < \frac{X}{30} < 0.05\}.$$

Example – continued

☞ First standardize X, then apply the normal approximation.

$$\begin{aligned}
\mathbf{P}\{-0.05 < \frac{X}{30} < 0.05\} &= \mathbf{P}\{-1.5 < X < 1.5\} \\
&= \mathbf{P}\{\frac{-1.5}{\sqrt{45}} < \frac{X-0}{\sqrt{45}} < \frac{1.5}{\sqrt{45}}\} \\
&\approx \Phi(\frac{1.5}{\sqrt{45}}) - \Phi(\frac{-1.5}{\sqrt{45}}) \\
&\approx \Phi(0.22) - \Phi(-0.22) = 2 \cdot \Phi(0.22) - 1 \\
&\approx 2(0.5871) - 1 \\
&= 0.1742.
\end{aligned}$$

☞ Thus, there is a 17.4% chance that the student's assignment average is accurate to within 0.05.

Example – continued. What is the probability that no error is made? That is, approximate

$$\mathbf{P}\{X = 0\} \qquad \text{where } X = X_1 + \ldots + X_{30}.$$

☞ Apply CLT, using continuity correction, since X is a discrete random variable and we want to compute the probability at a possible value.

$$\begin{aligned}
\mathbf{P}\{X = 0\} &= \mathbf{P}\{-0.5 \leq X \leq 0.5\} \\
&= \mathbf{P}\{\frac{-0.5-0}{\sqrt{45}} \leq \frac{X-0}{\sqrt{45}} \leq \frac{0.5-0}{\sqrt{45}}\} \\
&\approx \Phi(\frac{0.5}{\sqrt{45}}) - \Phi(\frac{-0.5}{\sqrt{45}}) \\
&\approx 2 \cdot \Phi(0.07) - 1 \\
&\approx 2(0.5279) - 1 = 0.0558
\end{aligned}$$

☞ There is only about a 5.6% that the recorded grade is correct.

Example

Example

Based on data of similar bridges, the span of a certain bridge can withstand a load, without structural damage, that is normally distributed with mean 400 and standard deviation 40 (in units of 1000 pounds). Suppose the weight of a car is a random variable with mean 3 and standard deviation 0.3 (in units of 1000 pounds).

Approximately how many cars would have to be on the bridge span for the probability of structural damage to exceed 10%?

Example – continued

☞ Let $X_1, X_2, \ldots$ be random variables denoting the weight of each car on the bridge, and let $S_n = X_1 + X_2 + \ldots + X_n$. Let Y denote the load the bridge can withstand. Then (where units are in 100 pounds)

$$\begin{aligned} E[X_i] &= 3 & SD(X_i) &= 0.3 \\ E[S_n] &= 3n & SD(S_n) &= 0.3\sqrt{n} \\ E[Y] &= 400 & SD(Y) &= 40 \end{aligned}$$

☞ We want to find n so that the probability

$$\mathbf{P}\{Y \le S_n\} \ge 0.1 \quad \text{equivalently} \quad \mathbf{P}\{0 \le S_n - Y\} \ge 0.1$$

Assume S_n and Y are independent, so

$$\mu_n = E[S_n - Y] = 3n - 400 \qquad \sigma_n = SD(S_n - Y) = \sqrt{(0.3)^2 n + (40)^2}$$

Example – continued

☞ First standardize, then apply the normal approximation:

$$\begin{aligned} 0.1 &\le \mathbf{P}\{0 \le S_n - Y\} \\ &= \mathbf{P}\{\frac{0 - \mu_n}{\sigma_n} \le \frac{S_n - Y - \mu_n}{\sigma_n}\} \\ &\approx 1 - \Phi(-\frac{\mu_n}{\sigma_n}) = \Phi(\frac{\mu_n}{\sigma_n}) \end{aligned}$$

Since $\Phi(x)$ increases as x increases, it is sufficient to have

$$\frac{\mu_n}{\sigma_n} \ge -1.29$$

or equivalently

$$\frac{3n - 400}{\sqrt{(0.3)^2 n + (40)^2}} \ge -1.29$$

Example – continued

$$\frac{3n - 400}{\sqrt{(0.3)^2 n + (40)^2}} \ge -1.29$$

☞ This reduces to a quadratic equation:

$$9n^2 - 2400n + 157,337 \le 0$$

The smallest value of n satisfying this equation is $n = 117$.

☞ If there are 117 or more cars on the bridge, then there is a greater than 10% chance the load on the bridge is greater then it can withstand without damage.