Section 8 – Solutions

Review

1) Multivariate: Discrete to Continuous:

	Discrete	Continuous
Joint PMF/PDF	$p_{X,Y}(x,y) = \mathbb{P}(X = x, Y = y)$	$f_{X,Y}(x,y) \neq \mathbb{P}(X = x, Y = y)$
Joint range/support $\Omega_{X,Y}$	$\{(x,y) \in \Omega_X \times \Omega_Y : p_{X,Y}(x,y) > 0\}$	$\{(x,y) \in \Omega_X \times \Omega_Y : f_{X,Y}(x,y) > 0\}$
Joint CDF	$F_{X,Y}(x,y) = \sum_{t \leqslant x, s \leqslant y} p_{X,Y}(t,s)$	$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(t,s)\, ds dt$
Normalization	$\sum_{x,y} p_{X,Y}(x,y) = 1$	$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx dy = 1$
Marginal PMF/PDF	$p_X(x) = \sum_y p_{X,Y}(x,y)$	$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy$
Expectation	$\mathbb{E}[g(X,Y)] = \sum_{x,y} g(x,y) p_{X,Y}(x,y)$	$\mathbb{E}[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y) dx dy$
Independence must have	$\forall x, y, p_{X,Y}(x,y) = p_X(x) p_Y(y)$ $\Omega_{X,Y} = \Omega_X \times \Omega_Y$	$\forall x, y, f_{X,Y}(x,y) = f_X(x) f_Y(y)$ $\Omega_{X,Y} = \Omega_X \times \Omega_Y$
Conditional PMF/PDF	$p_{X\|Y}(x\|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$	$f_{X\|Y}(x\|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$
Conditional Expectation	$\mathbb{E}[X\|Y = y] = \sum_x x \cdot p_{X\|Y}(x\|y)$	$\mathbb{E}[X\|Y = y] = \int_{-\infty}^{\infty} x f_{X\|Y}(x\|y) dx$

2) Normal (Gaussian, "bell curve"): $X \sim \mathcal{N}(\mu, \sigma^2)$ iff X has the following probability density function:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}, \quad x \in \mathbb{R}$$

$\mathbb{E}[X] = \mu$ and $\mathrm{Var}(X) = \sigma^2$. The "standard normal" random variable is typically denoted Z and has mean 0 and variance 1: if $X \sim \mathcal{N}(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$. The CDF has no closed form, but we denote the CDF of the standard normal as $\Phi(z) = F_Z(z) = \mathbb{P}(Z \leqslant z)$. Note from symmetry of the probability density function about $z = 0$ that: $\Phi(-z) = 1 - \Phi(z)$.

3) Central Limit Theorem (CLT): Let $X_1, \ldots, X_n$ be iid random variables with $\mathbb{E}[X_i] = \mu$ and $Var(X_i) = \sigma^2$.

Let $X = \sum_{i=1}^{n} X_i$, which has $\mathbb{E}[X] = n\mu$ and $Var(X) = n\sigma^2$. Let $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, which has $\mathbb{E}[\overline{X}] = \mu$ and $Var(\overline{X}) = \frac{\sigma^2}{n}$. $\overline{X}$ is called the *sample mean*. Then, as $n \to \infty$, $\overline{X}$ approaches the normal distribution $\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$. Standardizing, this is equivalent to $Y = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ approaching $\mathcal{N}(0,1)$. Similarly, as $n \to \infty$, X approaches $\mathcal{N}(n\mu, n\sigma^2)$ and $Y' = \frac{X-n\mu}{\sigma\sqrt{n}}$ approaches $\mathcal{N}(0,1)$.

It is no surprise that $\overline{X}$ has mean μ and variance σ^2/n – this can be done with simple calculations. The importance of the CLT is that, for large n, regardless of what distribution X_i comes from, $\overline{X}$ is *approximately normally distributed with mean* μ *and variance* σ^2/n. Don't forget the continuity correction, only when $X_1, \ldots, X_n$ are discrete random variables.

Here is the **Standard normal table.**

4) Law of Total Probability (Continuous): A is an event, and X is a continuous random variable with density function $f_X(x)$.

$$\mathbb{P}(A) = \int_{-\infty}^{\infty} \mathbb{P}(A \mid X = x) f_X(x) dx$$

Task 1 – Joint PMF's

Suppose X and Y have the following joint PMF:

X/Y	1	2	3
0	0	0.2	0.1
1	0.3	0	0.4

a) Identify the range of X (Ω_X), the range of Y (Ω_Y), and their joint range ($\Omega_{X,Y}$).

$\Omega_X = \{0, 1\}$, $\Omega_Y = \{1, 2, 3\}$, and $\Omega_{X,Y} = \{(0, 2), (0, 3), (1, 1), (1, 3)\}$

b) Find the marginal PMF for X, $p_X(x)$ for $x \in \Omega_X$.

Note that $\Omega_X = \{0, 1\}$.

$$p_X(0) = \sum_y p_{X,Y}(0, y) = 0 + 0.2 + 0.1 = 0.3$$

$$p_X(1) = 1 - p_X(0) = 0.7$$

c) Find the marginal PMF for Y, $p_Y(y)$ for $y \in \Omega_Y$.

Note that $\Omega_Y = \{1, 2, 3\}$.

$$p_Y(1) = \sum_x p_{X,Y}(x, 1) = 0 + 0.3 = 0.3$$

$$p_Y(2) = \sum_x p_{X,Y}(x, 2) = 0.2 + 0 = 0.2$$

$$p_Y(3) = \sum_x p_{X,Y}(x, 3) = 0.1 + 0.4 = 0.5$$

d) Are X and Y independent? Why or why not?

X and Y are not independent. Recall that a *necessary* condition for X and Y to be independent is that $\Omega_{X,Y} = \Omega_X \times \Omega_Y$. The joint range $\Omega_{X,Y}$ does not satisfy this criteria, so it cannot be independent.

e) Find $\mathbb{E}[X^3 Y]$.

Note that $X^3 = X$ since X takes values in $\{0, 1\}$.

$$\mathbb{E}[X^3 Y] = \mathbb{E}[XY] = \sum_{(x,y)\in\Omega_{X,Y}} xy p_{X,Y}(x, y) = 1 \cdot 1 \cdot 0.3 + 1 \cdot 3 \cdot 0.4 = 1.5$$

Task 2 – Do You "Urn" to Learn More About Probability?

Suppose that 3 balls are chosen without replacement from an urn consisting of 5 white and 8 red balls. Let $X_i = 1$ if the i-th ball selected is white and let it be equal to 0 otherwise. Give the joint probability mass function of

a) X_1, X_2

Here is one way of defining the joint pmf of X_1, X_2

$$p_{X_1,X_2}(1,1) = \mathbb{P}\left(X_1 = 1\right)\mathbb{P}\left(X_2 = 1 \mid X_1 = 1\right) = \frac{5}{13}\cdot\frac{4}{12} = \frac{20}{156}$$

$$p_{X_1,X_2}(1,0) = \mathbb{P}\left(X_1 = 1\right)\mathbb{P}\left(X_2 = 0 \mid X_1 = 1\right) = \frac{5}{13}\cdot\frac{8}{12} = \frac{40}{156}$$

$$p_{X_1,X_2}(0,1) = \mathbb{P}\left(X_1 = 0\right)\mathbb{P}\left(X_2 = 1 \mid X_1 = 0\right) = \frac{8}{13}\cdot\frac{5}{12} = \frac{40}{156}$$

$$p_{X_1,X_2}(0,0) = \mathbb{P}\left(X_1 = 0\right)\mathbb{P}\left(X_2 = 0 \mid X_1 = 0\right) = \frac{8}{13}\cdot\frac{7}{12} = \frac{56}{156}$$

b) X_1, X_2, X_3

Instead of listing out all the individual probabilities, we could write a more compact formula for the pmf. In this problem, the denominator is always $P(13, k)$, where k is the number of random variables in the joint pmf. And the numerator is $P(5, i)$ times $P(8, j)$ where i and j are the number of 1s and 0s, respectively.

If we wish to compute $p_{X_1,X_2,X_3}(x_1, x_2, x_3)$, then the number of 1s (i.e., white balls) is $x_1+x_2+x_3$, and the number of 0s (i.e., red balls) is $(1 - x_1) + (1 - x_2) + (1 - x_3)$. Then, we can write the pmf as follows:

$$p_{X_1,X_2,X_3}(x_1, x_2, x_3) = \frac{10!}{13!}\cdot\frac{5!}{(5 - x_1 - x_2 - x_3)!}\cdot\frac{8!}{(5 + x_1 + x_2 + x_3)!}$$

Task 3 – Trinomial Distribution

A generalization of the Binomial model is when there is a sequence of n independent trials, but with three outcomes, where $\mathbb{P}(\text{outcome } i) = p_i$ for $i = 1, 2, 3$ and of course $p_1 + p_2 + p_3 = 1$. Let X_i be the number of times outcome i occurred for $i = 1, 2, 3$, where $X_1 + X_2 + X_3 = n$. Find the joint PMF $p_{X_1,X_2,X_3}(x_1, x_2, x_3)$ and specify its value for all $x_1, x_2, x_3 \in \mathbb{R}$.

Are X_1 and X_2 independent?

In a similar argument with the binomial PMF, we have

$$p_{X_1,X_2,X_3}(x_1, x_2, x_3) = \binom{n}{x_1}\binom{n - x_1}{x_2}\binom{n - x_1 - x_2}{x_3} p_1^{x_1} p_2^{x_2} p_3^{x_3}\ .$$

This may also be interpreted as multinomial coefficients (reference), and so we may rewrite as

$$p_{X_1,X_2,X_3}(x_1, x_2, x_3) = \binom{n}{x_1, x_2, x_3} p_1^{x_1} p_2^{x_2} p_3^{x_3} = \frac{n!}{x_1!x_2!x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3}\ ,$$

where $x_1 + x_2 + x_3 = n$ and are nonnegative integers.

X_1 and X_2 are not independent. For example $\mathbb{P}(X_1 = n) > 0$ and $\mathbb{P}(X_2 = n) > 0$, but $\mathbb{P}(X_1 = n, X_2 = n) = 0$. In other words, $\Omega_{X_1,X_2,X_3} \neq \Omega_{X_1} \times \Omega_{X_2} \times \Omega_{X_3}$, which is a necessary condition for independence.

Task 4 – Successes

Consider a sequence of independent Bernoulli trials, each of which is a success with probability p. Let X_1 be the number of failures preceding the first success, and let X_2 be the number of failures after the first success but preceding the second success. Find the joint pmf of X_1 and X_2. Write an expression for $\mathbb{E}[\sqrt{X_1 X_2}]$. You can leave your answer in the form of a sum.

In order for X_1 to take on a particular value, say x_1, it must have x_1 failures until the first success, i.e., the next trial is a success. To that end, for X_1 and X_2 to take on two particular values x_1 and x_2, there must be x_1 failures followed by one success, and then x_2 failures followed by one success. Since the Bernoulli trials are independent, the joint pmf is

$$p_{X_1,X_2}(x_1,x_2) = (1-p)^{x_1}p \cdot (1-p)^{x_2}p = (1-p)^{x_1+x_2}p^2$$

for $(x_1,x_2) \in \Omega_{X_1,X_2} = \{0,1,2,\dots\} \times \{0,1,2,\dots\}$. By the definition of expectation and LOTUS,

$$\mathbb{E}[\sqrt{X_1 X_2}] = \sum_{(x_1,x_2)\in\Omega_{X_1,X_2}} \sqrt{x_1 x_2} \cdot p_{X_1,X_2}(x_1,x_2) = \sum_{(x_1,x_2)\in\Omega_{X_1,X_2}} \sqrt{x_1 x_2} \cdot (1-p)^{x_1+x_2}p^2 \,.$$

Task 5 – Who fails first?

Here's a question that commonly comes up in industry, but isn't immediately obvious. You have a disk with probability p_1 of failing each day. You have a CPU which independently has probability p_2 of failing each day. What is the probability that your disk fails *before* your CPU?

a) Compute the probability by summing over the relevant part of the probability space.

We model the problem by considering two Geometric random variables and deriving the probability that one is smaller than the other. Let $X_1 \sim \text{Geometric}(p_1)$. Let $X_2 \sim \text{Geometric}(p_2)$. Assume X_1 and X_2 are independent. We want $\mathbb{P}(X_1 < X_2)$.

$$\begin{aligned}
\mathbb{P}(X_1 < X_2) &= \sum_{k=1}^{\infty} \sum_{k_2=k+1}^{\infty} p_{X_1,X_2}(k,k_2) \\
&= \sum_{k=1}^{\infty} \sum_{k_2=k+1}^{\infty} p_{X_1}(k) \cdot p_{X_2}(k_2) && \text{(by independence)} \\
&= \sum_{k=1}^{\infty} \sum_{k_2=k+1}^{\infty} (1-p_1)^{k-1}p_1 \cdot (1-p_2)^{k_2-1}p_2 \\
&= \sum_{k=1}^{\infty} (1-p_1)^{k-1}p_1 \sum_{k_2=k+1}^{\infty} (1-p_2)^{k_2-1}p_2 \\
&= \sum_{k=1}^{\infty} (1-p_1)^{k-1}p_1(1-p_2)^k \sum_{k_2=1}^{\infty} (1-p_2)^{k_2-1}p_2 \\
&= \sum_{k=1}^{\infty} (1-p_1)^{k-1}p_1(1-p_2)^k \cdot 1 \\
&= p_1(1-p_2) \sum_{k=1}^{\infty} [(1-p_2)(1-p_1)]^{k-1} \\
&= \frac{p_1(1-p_2)}{1-(1-p_2)(1-p_1)}.
\end{aligned}$$

b) Try to provide an intuitive reason for the answer.

Think about X_1 and X_2 in terms of coin flips. Notice that all the flips are irrelevant until the final flip, since before the final flip, both the X_1 coin and the X_2 coin only yield tails. $\mathbb{P}(X_1 < X_2)$ is the probability that on the final flip, where by definition at least one coin comes up heads, it is the case that the X_1 coin is heads and the X_2 coin is tails. So we're looking for the probability that the X_1 coin produces a heads and the X_2 coin produces a tails, conditioned on the fact that they're not both tails, which is derived as:

$$\begin{aligned}\mathbb{P}(\text{Coin 1} = H \text{ and Coin 2} = T \mid \text{not both } T) &= \frac{\mathbb{P}(\text{Coin 1} = H \text{ and Coin 2} = T)}{\mathbb{P}(\text{not both } T)} \\ &= \frac{p_1(1-p_2)}{1-(1-p_2)(1-p_1)}.\end{aligned}$$

Another way to approach this problem is to use conditioning. Recall that in computing the probability of an event, we saw in Chapter 2 that it is often useful to condition on other events. We can use this same idea in computing probabilities involving random variables, because $X = k$ and $Y = y$ are just events.

c) Recompute the probability using the law of total probability, conditioning on the value of X_1.

Again, let $X_1 \sim \text{Geometric}(p_1)$ and $X_2 \sim \text{Geometric}(p_2)$, where X_1 and X_2 are independent. Then

$$\begin{aligned}\mathbb{P}(X_1 < X_2) &= \sum_{k=1}^{\infty} \mathbb{P}(X_1 < X_2 \mid X_1 = k) \cdot \mathbb{P}(X_1 = k) \\ &= \sum_{k=1}^{\infty} \mathbb{P}(k < X_2 \mid X_1 = k) \cdot \mathbb{P}(X_1 = k) \\ &= \sum_{k=1}^{\infty} \mathbb{P}(X_2 > k) \cdot \mathbb{P}(X_1 = k) && \text{(by independence)} \\ &= \sum_{k=1}^{\infty} (1-p_2)^k \cdot (1-p_1)^{k-1} \cdot p_1 \\ &= p_1(1-p_2) \sum_{k=1}^{\infty} [(1-p_2)(1-p_1)]^{k-1} \\ &= \frac{p_1(1-p_2)}{1-(1-p_2)(1-p_1)}.\end{aligned}$$

Task 6 – Continuous joint density

The joint density of X and Y is given by

$$f_{X,Y}(x,y) = \begin{cases} xe^{-(x+y)} & x > 0, y > 0 \\ 0 & \text{otherwise.} \end{cases}$$

and the joint density of W and V is given by

$$f_{W,V}(w,v) = \begin{cases} 2 & 0 < w < v, 0 < v < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Are X and Y independent? Are W and V independent?

For two random variables X, Y to be independent, we must have $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ for all $x \in \Omega_X,\ y \in \Omega_Y$. Let's start with X and Y by finding their marginal PDFs. By definition, and using the fact that the joint PDF is 0 outside of $y > 0$, we get:

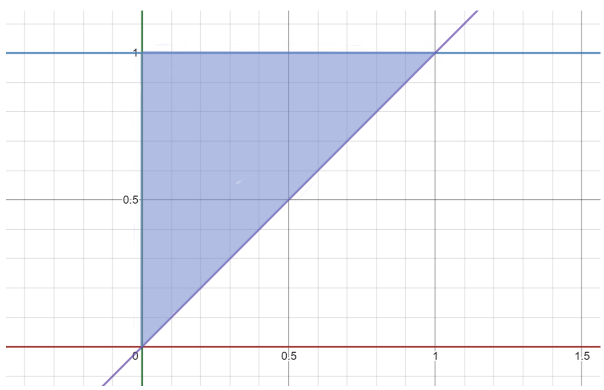
$$f_X(x) = \int_0^\infty xe^{-(x+y)}dy = e^{-x}x$$

We do the same to get the PDF of Y, again over the range $x > 0$:

$$f_Y(y) = \int_0^\infty xe^{-(x+y)}dx = e^{-y}$$

Since $e^{-x}x \cdot e^{-y} = xe^{-x-y} = xe^{-(x+y)}$ for all $x, y > 0$, X and Y are independent.

We can see that W and V are not independent simply by observing that $\Omega_W = (0, 1)$ and $\Omega_V = (0, 1)$, but $\Omega_{W,V}$ is not equal to their Cartesian product. Specifically, looking at their range of $f_{W,V}(w, v)$. Graphing it with w as the "x-axis" and v as the "y-axis", we see that :



The shaded area is where the joint pdf is strictly positive. Looking at it, we can see that it is not rectangular, and therefore it is not the case that $\Omega_{W,V} = \Omega_W \times \Omega_V$. Remember, the joint range being the Cartesian product of the marginal ranges is not sufficient for independence, but it is *necessary*. Therefore, this is enough to show that they are not independent.

Task 7 – Grades and homework turn-in time

Suppose we're currently trying to find a relationship between the time a student turns in their homework and the grade that they receive on the respective homework. Let T denote the amount of time *prior* to the deadline that the homework is submitted. We have observed that no student submits the homework more than 2 days earlier than the deadline, and also no student submits their assignment late, so $0 \leqslant T \leqslant 2$. Now let G be a random variable, indicating the percentage that the student receives on the homework assignment, that is, $0 \leqslant G \leqslant 1$. Suppose G and T are continuous random variables, and their joint pdf is given by

$$f_{G,T}(g, t) = \begin{cases} \frac{9}{10}g^2t + \frac{1}{5} & \text{when } 0 \leqslant g \leqslant 1 \text{ and } 0 \leqslant t \leqslant 2 \\ 0 & \text{otherwise .} \end{cases}$$

For both parts, round your solution to three decimal places.

a) What is the probability that a randomly selected student gets a grade above 50% on the homework?

We are looking for $\mathbb{P}(G > 0.5)$. To do this, we must first compute the marginal density function $f_G(g)$. Applying by definition,

$$\begin{aligned} f_G(g) &= \int_{-\infty}^{\infty} f_{G,T}(g,t)\,dt \\ &= \int_0^2 f_{G,T}(g,t)\,dt = \int_0^2 \frac{9}{10}g^2 t + \frac{1}{5}\,dt = \left(\frac{9}{10}\frac{1}{2}t^2 g^2 + \frac{1}{5}t\right)\Big|_0^2 = \frac{9}{5}g^2 + \frac{2}{5}\,. \end{aligned}$$

Then

$$\mathbb{P}(G > 0.5) = \int_{0.5}^{\infty} f_G(g)\,dg = \int_{0.5}^{1} \frac{9}{5}g^2 + \frac{2}{5}\,dg = \frac{29}{40} = 0.725\,.$$

b) What is the probability that a student gets a grade above 50%, given that the student submitted less than a day before the deadline?

We are looking for

$$\mathbb{P}(G > 0.5 \mid T < 1) = \frac{\mathbb{P}(G > 0.5 \cap T < 1)}{\mathbb{P}(T < 1)}\,,$$

which follows by the definition of conditional probability. The numerator can be computed using the joint pdf. However, the denominator needs us to calculate the marginal pdf. We can follow a similar approach to the previous part and get

$$f_T(t) = \int_0^1 f_{G,T}(g,t)\,dg = \int_0^1 \frac{9}{10}g^2 t + \frac{1}{5}\,dg = \frac{3}{10}t + \frac{1}{5}\,.$$

Thus,

$$\mathbb{P}(G > 0.5 \mid T < 1) = \frac{\int_{0.5}^1 \int_0^1 f_{G,T}(g,t)\,dt\,dg}{\int_0^1 f_T(t)\,dt} = \frac{\int_{0.5}^1 \int_0^1 \frac{9}{10}g^2 t + \frac{1}{5}\,dt\,dg}{\int_0^1 \frac{3}{10}t + \frac{1}{5}\,dt} \approx 0.661\,.$$

Task 8 – Confidence Intervals

Suppose that $X_1, \ldots, X_n$ are i.i.d. samples from a normal distribution with unknown mean μ and variance 36. How big does n need to be so that $\mathbb{E}[\overline{X}] = \mu$ is in

$$[\overline{X} - 0.11, \overline{X} + 0.11]$$

with probability at least 0.97? Recall that

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

You may use the fact that $\Phi^{-1}(0.985) = 2.17$.

Our goal is to find n such that μ lies within 0.11 of $\bar{X}$ 97% of the time. This is equivalent to finding n such that the probability that μ lies outside the range is less than 3%.

$$\mathbb{P}(|\bar{X} - \mu| > 0.11) \leqslant 0.03$$

Let us define $Z = \frac{\bar{X}-\mu}{\sigma}$. We can solve for σ by using the Properties of Variance. Since

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

we can say that

$$\text{Var}(\bar{X}) = \text{Var}(\frac{1}{n}\sum_{i=1}^{n} X_i)$$

Using the Properties of Variance and the fact that X_i's are i.i.d., $\text{Var}(\bar{X}) = \frac{1}{n^2} \cdot n \cdot 36 = \frac{36}{n}$, so $\sigma = \frac{6}{\sqrt{n}}$.

$$\begin{aligned}
\mathbb{P}(|\bar{X} - \mu| > 0.11) &\leqslant 0.03 \\
\mathbb{P}(|Z| \cdot \sigma > 0.11) &\leqslant 0.03 && \text{[Definition of } Z\text{]} \\
\mathbb{P}\left(|Z| > \frac{0.11}{6}\sqrt{n}\right) &\leqslant 0.03 \\
\mathbb{P}\left(Z < -\frac{0.11}{6}\sqrt{n}\right) &\leqslant 0.015 && \text{[Symmetry of Normal Dist.]} \\
\Phi\left(-\frac{0.11}{6}\sqrt{n}\right) &\leqslant 0.015 && \text{[CDF of Standard Norm.]} \\
-\frac{0.11}{6}\sqrt{n} &\leqslant -\Phi^{-1}(0.985) \\
\sqrt{n} &\geqslant \frac{6 \cdot \Phi^{-1}(0.985)}{0.11} \\
n &\geqslant \left(\frac{6 \cdot \Phi^{-1}(0.985)}{0.11}\right)^2 \\
&\approx 14009.95
\end{aligned}$$

Then n must be at least 14010.