ECS 315: Probability and Random Processes **2016/1**

HW Solution 13 — Due: Dec 6, 5 PM

Lecturer: Prapun Suksompong, Ph.D.

Problem 1. A random variable X is a Gaussian random variable if its pdf is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2},$$

for some constant m and positive number σ. Furthermore, when a Gaussian random variable has $m = 0$ and $\sigma = 1$, we say that it is a standard Gaussian random variable. There is no closed-form expression for the cdf of the standard Gaussian random variable. The cdf itself is denoted by Φ and its values (or its complementary values $Q(\cdot) = 1 - \Phi(\cdot)$) are traditionally provided by a table.

Suppose Z is a standard Gaussian random variable.

(a) Use the Φ table to find the following probabilities:

(i) $P[Z < 1.52]$

(ii) $P[Z < -1.52]$

(iii) $P[Z > 1.52]$

(iv) $P[Z > -1.52]$

(v) $P[-1.36 < Z < 1.52]$

(b) Use the Φ table to find the value of c that satisfies each of the following relation.

(i) $P[Z > c] = 0.14$

(ii) $P[-c < Z < c] = 0.95$

Solution:

(a)

(i) $P[Z < 1.52] = \Phi(1.52) = \boxed{0.9357.}$

(ii) $P[Z < -1.52] = \Phi(-1.52) = 1 - \Phi(1.52) = 1 - 0.9357 = \boxed{0.0643.}$

(iii) $P[Z > 1.52] = 1 - P[Z < 1.52] = 1 - \Phi(1.52) = 1 - 0.9357 = \boxed{0.0643.}$

(iv) It is straightforward to see that the area of $P[Z > -1.52]$ is the same as $P[Z < 1.52] = \Phi(1.52)$. So, $P[Z > -1.52] = \boxed{0.9357.}$
Alternatively, $P[Z > -1.52] = 1 - P[Z \le -1.52] = 1 - \Phi(-1.52) = 1 - (1 - \Phi(1.52)) = \Phi(1.52)$.

(v) $P[-1.36 < Z < 1.52] = \Phi(1.52) - \Phi(-1.36) = \Phi(1.52) - (1 - \Phi(1.36)) = \Phi(1.52) + \Phi(1.36) - 1 = 0.9357 + 0.9131 - 1 = \boxed{0.8488.}$

(b)

(i) $P[Z > c] = 1 - P[Z \leq c] = 1 - \Phi(c)$. So, we need $1 - \Phi(c) = 0.14$ or $\Phi(c) = 1 - 0.14 = 0.86$. In the Φ table, we do not have exactly 0.86, but we have 0.8599 and 0.8621. Because 0.86 is closer to 0.8599, we answer the value of c whose $\phi(c) = 0.8599$. Therefore, $c \approx \boxed{1.08.}$

(ii) $P[-c < Z < c] = \Phi(c) - \Phi(-c) = \Phi(c) - (1 - \Phi(c)) = 2\Phi(c) - 1$. So, we need $2\Phi(c) - 1 = 0.95$ or $\Phi(c) = 0.975$. From the Φ table, we have $c \approx \boxed{1.96.}$

Problem 2. The peak temperature T, as measured in degrees Fahrenheit, on a July day in New Jersey is a $\mathcal{N}(85, 100)$ random variable.

Remark: Do not forget that, for our class, the second parameter in $\mathcal{N}(\cdot, \cdot)$ is the variance (not the standard deviation).

(a) Express the cdf of T in terms of the Φ function.

(b) Express each of the following probabilities in terms of the Φ function(s). Make sure that the arguments of the Φ functions are positive. (Positivity is required so that we can directly use the Φ/Q tables to evaluate the probabilities.)

(i) $P[T > 100]$

(ii) $P[T < 60]$

(iii) $P[70 \leq T \leq 100]$

(c) Express each of the probabilities in part (b) in terms of the Q function(s). Again, make sure that the arguments of the Q functions are positive.

(d) Evaluate each of the probabilities in part (b) using the Φ/Q tables.

(e) Observe that the Φ table ("Table 4" from the lecture) stops at $z = 2.99$ and the Q table ("Table 5" from the lecture) starts at $z = 3.00$. Why is it better to give a table for $Q(z)$ instead of $\Phi(z)$ when z is large?

Solution:

(a) Recall that when $X \sim \mathcal{N}(m, \sigma^2)$, $F_X(x) = \Phi\left(\frac{x-m}{\sigma}\right)$. Here, $T \sim \mathcal{N}(85, 10^2)$. Therefore,

$$F_T(t) = \boxed{\Phi\left(\frac{t-85}{10}\right).}$$

(b)

(i) $P[T > 100] = 1 - P[T \le 100] = 1 - F_T(100) = 1 - \Phi\left(\frac{100-85}{10}\right) = 1 - \Phi(1.5)$

(ii) $P[T < 60] = P[T \le 60]$ because T is a continuous random variable and hence $P[T = 60] = 0$. Now, $P[T \le 60] = F_T(60) = \Phi\left(\frac{60-85}{10}\right) = \Phi(-2.5) = \boxed{1 - \Phi(2.5).}$ Note that, for the last equality, we use the fact that $\Phi(-z) = 1 - \Phi(z)$.

(iii)

$$\begin{aligned} P[70 \le T \le 100] &= F_T(100) - F_T(70) = \Phi\left(\frac{100-85}{10}\right) - \Phi\left(\frac{70-85}{10}\right) \\ &= \Phi(1.5) - \Phi(-1.5) = \Phi(1.5) - (1 - \Phi(1.5)) = \boxed{2\Phi(1.5) - 1.} \end{aligned}$$

(c) In this question, we use the fact that $Q(x) = 1 - \Phi(x)$.

(i) $1 - \Phi(1.5) = \boxed{Q(1.5).}$

(ii) $1 - \Phi(2.5) = \boxed{Q(2.5).}$

(iii) $2\Phi(1.5) - 1 = 2(1 - Q(1.5)) - 1 = 2 - 2Q(1.5) - 1 = \boxed{1 - 2Q(1.5).}$

(d)

(i) $1 - \Phi(1.5) = 1 - 0.9332 = \boxed{0.0668.}$

(ii) $1 - \Phi(2.5) = 1 - 0.99379 = \boxed{0.0062.}$

(iii) $2\Phi(1.5) - 1 = 2(0.9332) - 1 = \boxed{0.8664.}$

(e) When z is large, $\Phi(z)$ will start with 0.999... The first few significant digits will all be the same and hence not quite useful to be there.

Problem 3. Suppose that the time to failure (in hours) of fans in a personal computer can be modeled by an exponential distribution with $\lambda = 0.0003$.

(a) What proportion of the fans will last at least 10,000 hours?

(b) What proportion of the fans will last at most 7000 hours?

[Montgomery and Runger, 2010, Q4-97]

Solution: See handwritten solution.

HW13 Q3: Exponential RV

Friday, November 21, 2014 9:02 AM

Let T be the time to failure (in hours)

We know that $T \sim \mathcal{E}(\lambda)$ where $\lambda = 3 \times 10^{-4}$.

Therefore,

$$f_T(t) = \begin{cases} \lambda e^{-\lambda t}, & t > 0, \\ 0, & \text{otherwise}. \end{cases}$$

(a) Here, we want to find $P[T > 10^4]$.

We will first provide the general formula for $P[T > t]$.

For $T \sim \mathcal{E}(\lambda)$ and $t > 0$,

$$P[T > t] = \int_t^\infty f_T(\tau)\,d\tau = \int_t^\infty \lambda e^{-\lambda\tau}\,d\tau = -e^{-\lambda\tau}\Big|_t^\infty = e^{-\lambda t}$$

Therefore, $P[T > 10^4] = e^{-3\times10^{-4}\times10^4} = e^{-3} \approx 0.0498$

(b) $P[T \leq 7000] = 1 - P[T > 7000] = 1 - e^{-3\times10^{-4}\times7000} = 1 - e^{-2.1} \approx 0.8775$

Remark: In class, we have already shown that for $T \sim \mathcal{E}(\lambda)$,

$$F_T(t) = \begin{cases} 1 - e^{-\lambda t}, & t > 0, \\ 0, & \text{otherwise}, \end{cases}$$

and

$$P[T > t] = \begin{cases} e^{-\lambda t}, & t > 0, \\ 1, & \text{otherwise}. \end{cases}$$

These formula can be applied here directly as well.

Problem 4. Consider each random variable X defined below. Let $Y = 1 + 2X$. (i) Find and sketch the pdf of Y and (ii) Does Y belong to any of the (popular) families discussed in class? If so, state the name of the family and find the corresponding parameters.

(a) $X \sim \mathcal{U}(0, 1)$

(b) $X \sim \mathcal{E}(1)$

(c) $X \sim \mathcal{N}(0, 1)$

Solution: See handwritten solution

Problem 5. Consider each random variable X defined below. Let $Y = 1 - 2X$. (i) Find and sketch the pdf of Y and (ii) Does Y belong to any of the (popular) families discussed in class? If so, state the name of the family and find the corresponding parameters.

(a) $X \sim \mathcal{U}(0, 1)$

(b) $X \sim \mathcal{E}(1)$

(c) $X \sim \mathcal{N}(0, 1)$

Solution: See handwritten solution

HW13 Q4 Affine Transformation

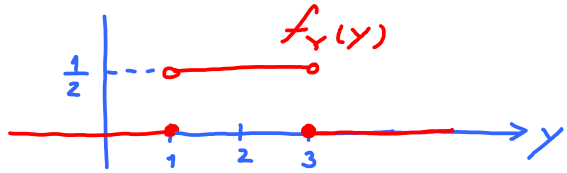
Tuesday, November 11, 2014 4:15 PM

We know that when $Y = aX + b$, we have $f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$.

Here, $Y = \underbrace{2}_{a}X + \underbrace{1}_{b}$. Therefore, $f_Y(y) = \frac{1}{2} f_X\left(\frac{y-1}{2}\right)$.

(a) $X \sim \mathcal{U}(0,1) \Rightarrow f_X(x) = \begin{cases} 1, & 0 < x < 1, \\ 0, & \text{otherwise.} \end{cases}$

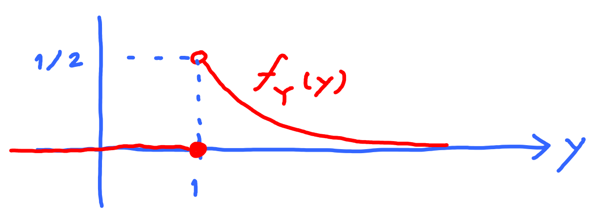
(i) Therefore, $f_Y(y) = \frac{1}{2} \times \begin{cases} 1, & 0 < \frac{y-1}{2} < 1 \\ 0, & \text{otherwise} \end{cases} = \begin{cases} 1/2, & 1 < y < 3, \\ 0, & \text{otherwise.} \end{cases}$



(ii) Yes. $Y \sim \mathcal{U}(1,3)$

(b) $X \sim \mathcal{E}(1) \Rightarrow f_X(x) = \begin{cases} 1 \times e^{-1 \times x}, & x > 0, \\ 0, & \text{otherwise} \end{cases} = \begin{cases} e^{-x}, & x > 0, \\ 0, & \text{otherwise.} \end{cases}$

(i) Therefore, $f_Y(y) = \frac{1}{2} \times \begin{cases} e^{-\left(\frac{y-1}{2}\right)}, & \frac{y-1}{2} > 0, \\ 0, & \text{otherwise} \end{cases} = \begin{cases} \frac{1}{2}\sqrt{e}\, e^{-y/2}, & y > 1, \\ 0, & \text{otherwise.} \end{cases}$



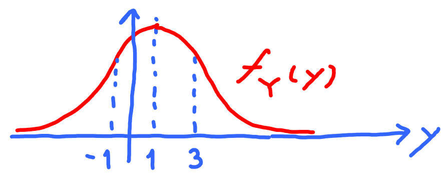
(ii) No. Although f_Y decays exponentially, the "exponential part" starts @ $y = 1$ (not @ $y = 0$). We may call this distribution a shifted exponential distribution. This distribution is quite useful for modeling output of a biological neuron with refractory period.

(c) We know that $X \sim \mathcal{N}(m, \sigma^2) \Rightarrow Y = \overset{2}{a}X + \overset{1}{b} \sim \mathcal{N}(am+b, a^2\sigma^2)$.

Plugging in $a=2$ and $b=1$, we have $Y \sim \mathcal{N}(2m+1, 4\sigma^2)$.

Here, $X \sim \mathcal{N}(\underset{m}{0}, \underset{\sigma^2}{1})$. So, $Y \sim \mathcal{N}(2\times 0+1, 4\times 1) = \mathcal{N}(1, 4)$ $\rightarrow \sigma = 2$

(i) $f_Y(y) = \dfrac{1}{\sqrt{2\pi}\,2} e^{-\frac{1}{2}\left(\frac{y-1}{2}\right)^2}$



(ii) Yes. $Y \sim \mathcal{N}(1, 4)$

HW13 Q5 Affine Transformation

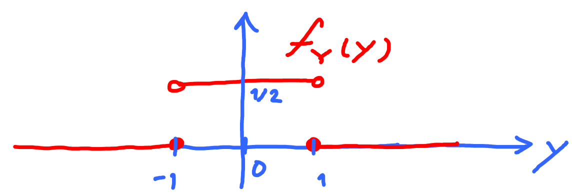
Tuesday, November 11, 2014 4:15 PM

We know that when $Y = aX+b$, we have $f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$.

Here, $Y = \underset{a}{-2}X + \underset{b}{1}$. Therefore, $f_Y(y) = \frac{1}{2} f_X\left(\frac{1-y}{2}\right)$.

(a) $X \sim \mathcal{U}(0,1) \Rightarrow f_X(x) = \begin{cases} 1, & 0<x<1, \\ 0, & \text{otherwise.} \end{cases}$

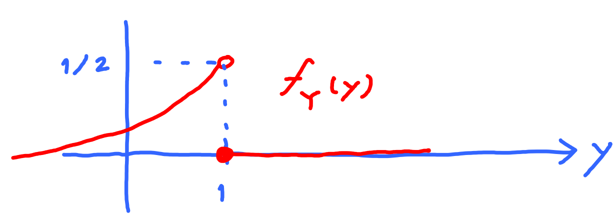
(i) Therefore, $f_Y(y) = \frac{1}{2} \times \begin{cases} 1, & 0<\frac{1-y}{2}<1 \\ 0, & \text{otherwise} \end{cases} = \begin{cases} 1/2, & -1<y<1, \\ 0, & \text{otherwise.} \end{cases}$



(ii) Yes. $Y \sim \mathcal{U}(-1,1)$

(b) $X \sim \mathcal{E}(1) \Rightarrow f_X(x) = \begin{cases} 1 \times e^{-1 \times x}, & x>0, \\ 0, & \text{otherwise} \end{cases} = \begin{cases} e^{-x}, & x>0, \\ 0, & \text{otherwise.} \end{cases}$

(i) Therefore, $f_Y(y) = \frac{1}{2} \times \begin{cases} e^{-\left(\frac{1-y}{2}\right)}, & \frac{1-y}{2}>0, \\ 0, & \text{otherwise} \end{cases} = \begin{cases} \frac{1}{2\sqrt{e}} e^{y/2}, & y<1, \\ 0, & \text{otherwise.} \end{cases}$



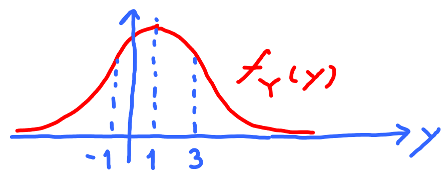
(ii) No. Although f_Y has exponential decay, its expression can not be rewritten in the form $f_Y(y) = \begin{cases} \lambda e^{-\lambda y}, & y>0, \\ 0, & \text{otherwise.} \end{cases}$

(c) We know that $X \sim \mathcal{N}(m, \sigma^2) \Rightarrow Y = \overset{-2}{a}X + \overset{1}{b} \sim \mathcal{N}(am+b, a^2\sigma^2)$.

Plugging in $a = -2$ and $b = 1$, we have $Y \sim \mathcal{N}(1-2m, 4\sigma^2)$

Here, $X \sim \mathcal{N}(\underset{m}{0}, \underset{\sigma^2}{1})$. So, $Y \sim \mathcal{N}(-2\times 0+1, 4\times 1) = \mathcal{N}(1, \underset{\sigma = 2}{4})$

(i) $f_Y(y) = \dfrac{1}{\sqrt{2\pi}\,2} e^{-\frac{1}{2}\left(\frac{y-1}{2}\right)^2}$



(ii) Yes. $Y \sim \mathcal{N}(1, 4)$

Problem 6. Let $X \sim \mathcal{E}(3)$.

(a) For each of the following function $g(x)$. Indicate whether the random variable $Y = g(X)$ is a continuous random variable.

(i) $g(x) = x^2$.

(ii) $g(x) = \begin{cases} 1, & x \geq 0, \\ 0, & x < 0. \end{cases}$.

(iii) $g(x) = \begin{cases} 4e^{-4x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$

(iv) $g(x) = \begin{cases} x, & x \leq 5, \\ 5, & x > 5. \end{cases}$

(b) Repeat part (a), but now check whether the random variable $Y = g(X)$ is a discrete random variable.

Solution:

(a)

(i) $\boxed{\text{YES}}$.

When $y < 0$, there is no x that satisfies $g(x) = y$. When $y = 0$, there is exactly one x ($x = 0$) that satisfies $g(x) = y$. When $y > 0$, there is exactly two x ($x = \pm\sqrt{y}$) that satisfies $g(x) = y$.

Therefore, for any y, there are at most countably many x values that satisfy $g(X) = y$. Because X is a continuous random variable, we conclude that Y is also a continuous random variable.

(ii) $\boxed{\text{NO}}$. An easy way to see this is that there can be only two values out of the function $g(\cdot)$: 0 or 1. So, $Y = g(X)$ is a discrete random variable.

Alternatively, consider $y = 1$. We see that any $x \geq 0$, can make $g(x) = 1$. Therefore, $P[Y = 1] = P[X \geq 0]$. For $X \sim \mathcal{E}(3)$, $P[X \geq 0] = 1 > 0$.

Because we found a y with $P[Y = y] > 0$. Y can not be a continuous random variable.

(iii) $\boxed{\text{YES}}$.

The plot of the function $g(x)$ may help you see the following facts: When $y > 4$ or $y < 0$, there is no x that gives $y = g(x)$. When $0 < y < 4$, there is exactly one x that satisfies $y = g(x)$. Because X is a continuous random variable, we can conclude that $P[Y = y]$ is 0 for $y \neq 0$.

When $y = 0$, any $x < 0$ would satisfy $g(x) = y$. So, $P[Y = 0] = P[X < 0]$. However, because $X \sim \mathcal{E}(3)$ is always positive. $P[X < 0] = 0$.

(iv) $\boxed{\text{NO}}$. Consider $y = 5$. We see that any $x \geq 5$, can make $g(x) = 5$. Therefore,

$$P[Y = 5] = P[X \geq 5].$$

For $X \sim \mathcal{E}(3)$,

$$P[X \geq 5] = \int_5^\infty 3e^{-3x} dx = e^{-15} > 0.$$

Because $P[Y = 5] > 0$, we conclude that Y can't be a continuous random variable.

(b) To check whether a random variable is discrete, we simply check whether it has a countable support. Also, if we have already checked that a random variable is continuous, then it can't also be discrete.

(i) $\boxed{\text{NO}}$. We checked before that it is a continuous random variable.

(ii) $\boxed{\text{YES}}$ as discussed in part (a).

(iii) $\boxed{\text{NO}}$. We checked before that it is a continuous random variable.

(iv) $\boxed{\text{NO}}$. Because X is positive, $Y = g(X)$ can be any positive number in the interval $(0, 5]$. The interval is uncountable. Therefore, Y is not discrete.

We have shown previously that Y is not a continuous random variable. Here. knowing that it is not discrete means that it is of the last type: mixed random variable.

Problem 7. Cholesterol is a fatty substance that is an important part of the outer lining (membrane) of cells in the body of animals. Its normal range for an adult is 120–240 mg/dl. The Food and Nutrition Institute of the Philippines found that the total cholesterol level for Filipino adults has a mean of 159.2 mg/dl and 84.1% of adults have a cholesterol level below 200 mg/dl. Suppose that the cholesterol level in the population is normally distributed.

(a) Determine the standard deviation of this distribution.

(b) What is the value of the cholesterol level that exceeds 90% of the population?

(c) An adult is at moderate risk if cholesterol level is more than one but less than two standard deviations above the mean. What percentage of the population is at moderate risk according to this criterion?

(d) An adult is thought to be at high risk if his cholesterol level is more than two standard deviations above the mean. What percentage of the population is at high risk?

Solution: See handwritten solution

HW13 Q7: Gaussian RV - Cholesterol

Thursday, November 13, 2014 3:49 PM

Let X be the cholesterol level of a randomly chosen adult.

It is given that $X \sim \mathcal{N}(m, \sigma^2)$ where $m = 159.2$ mg/dl.

We also know that $P[X < 200] = 0.841$.

(a) $\Phi\left(\frac{200-m}{\sigma}\right) = 0.841 \Rightarrow \frac{200-m}{\sigma} \approx 1 \Rightarrow \sigma \approx 200 - m \approx 200 - 159.2 \approx 40.8$ mg/dl

From the Φ table,

$\Phi(0.99) \approx 0.8389$

$\Phi(1) \approx 0.8413$

0.841 is closer to 0.8413 than 0.8389

(b) We find x such that $\Phi\left(\frac{x-m}{\sigma}\right) = 0.9$.

From the Φ table, $\Phi(1.28) \approx 0.8997$

$\Phi(1.29) \approx 0.9015$

0.9 is closer to 0.8997

$\Rightarrow \frac{x-m}{\sigma} \approx 1.28 \Rightarrow x \approx 1.28\underbrace{\sigma}_{\substack{40.8 \\ \text{(from (a))}}} + \underbrace{m}_{159.2} \approx 211.424$ mg/dl

(c) $P[m+\sigma < X < m+2\sigma] = F_X(m+2\sigma) - F_X(m+\sigma)$

$= \Phi\left(\frac{m+2\sigma-m}{\sigma}\right) - \Phi\left(\frac{m+\sigma-m}{\sigma}\right) = \Phi(2) - \Phi(1)$

$\approx 0.97725 - 0.8413 = 0.1359 \approx 13.59\%$

(d) $P[X > m+2\sigma] = 1 - F_X(m+2\sigma) = 1 - \Phi\left(\frac{m+2\sigma-m}{\sigma}\right) = 1 - \Phi(2)$

$\approx 1 - 0.97725 \approx 0.0228 = 2.28\%$

Problem 8 (Q3.5.6)**.** Solve this question using the Φ/Q table.

A professor pays 25 cents for each blackboard error made in lecture to the student who points out the error. In a career of n years filled with blackboard errors, the total amount in dollars paid can be approximated by a Gaussian random variable Y_n with expected value $40n$ and variance $100n$.

(a) What is the probability that Y_{20} exceeds 1000?

(b) How many years n must the professor teach in order that $P[Y_n > 1000] > 0.99$?

Solution: We are given[1] that $Y_n \sim \mathcal{N}(40n, 100n)$. Recall that when $X \sim \mathcal{N}(m, \sigma^2)$,

$$F_X(x) = \Phi\left(\frac{x-m}{\sigma}\right). \tag{13.1}$$

(a) Here $n = 20$. So, we have $Y_n \sim \mathcal{N}(40 \times 20, 100 \times 20) = \mathcal{N}(800, 2000)$. For this random variable $m = 800$ and $\sigma = \sqrt{2000}$.

We want to find $P[Y_{20} > 1000]$ which is the same as $1 - P[Y_20 \le 1000]$. Expressing this quantity using cdf, we have

$$P[Y_{20} > 1000] = 1 - F_{Y_{20}}(1000).$$

Apply (13.1) to get

$$P[Y_{20} > 1000] = 1 - \Phi\left(\frac{1000-800}{\sqrt{2000}}\right) = 1 - \Phi(4.472) \approx Q(4.47) \approx \boxed{3.91 \times 10^{-6}}.$$

(b) Here, the value of n is what we want. So, we will need to keep the formula in the general form. Again, from (13.1), for $Y_n \sim \mathcal{N}(40n, 100n)$, we have

$$P[Y_n > 1000] = 1 - F_{Y_n}(1000) = 1 - \Phi\left(\frac{1000-40n}{10\sqrt{n}}\right) = 1 - \Phi\left(\frac{100-4n}{\sqrt{n}}\right).$$

To find the value of n such that $P[Y_n > 1000] > 0.99$, we will first find the value of z which make

$$1 - \Phi(z) > 0.99. \tag{13.2}$$

At this point, we may try to solve for the value of Z by noting that (13.2) is the same as

$$\Phi(z) < 0.01. \tag{13.3}$$

[1]Note that the expected value and the variance in this question are proportional to n. This naturally occurs when we consider the sum of i.i.d. random variables. The approximation by Gaussian random variable is a result of the central limit theorem (CLT).

Unfortunately, the tables that we have start with $\Phi(0) = 0.5$ and increase to something close to 1 when the argument of the Φ function is large. This means we can't directly find 0.01 in the table. Of course, 0.99 is in there and therefore we will need to solve (13.2) via another approach.

To do this, we use another property of the Φ function. Recall that $1 - \Phi(z) = \Phi(-z)$. Therefore, (13.2) is the same as

$$\Phi\left(-z\right) > 0.99. \tag{13.4}$$

From our table, we can then conclude that (13.3) (which is the same as (13.4)) will happen when $-z > 2.33$. (If you have `MATLAB`, then you can get a more accurate answer of 2.3263.)

Now, plugging in $z = \frac{100-4n}{\sqrt{n}}$, we have $\frac{4n-100}{\sqrt{n}} > 2.33$. To solve for n, we first let $x = \sqrt{n}$. In which case, we have $\frac{4x^2-100}{x} > 2.33$ or, equivalently, $4x^2 - 2.33x - 100 > 0$. The two roots are $x = -4.717$ and $x > 5.3$. So, We need $x < -4.717$ or $x > 5.3$. Note that $x = \sqrt{n}$ and therefore can not be negative. So, we only have one case; that is, we need $x > 5.3$. Because $n = x^2$, we then conclude that we need $\boxed{n > 28.1 \text{ years}}$.