Continuous random variables

So far everything we have studied have been *discrete*. Discrete distributions, discrete random variables. However, many interesting things are *continuous*, so we need a new model to handle these things.

Definition. A **probability density function (pdf)** for a continuous random variable X, is a function $f : \mathbb{R} \to \mathbb{R}$ such that

$$\int_{-\infty}^{\infty} f(x)\,dx = 1 \text{ and } f(x) \geq 0 \quad \forall x$$

We must also have that $\mathbb{P}(a \leq X \leq b) = \int_a^b f(x)\,dx$

Note that $\mathbb{P}(X = a) = \int_a^a f(x)\,dx = 0 \neq f(a)$

Example

Suppose

$$f(x) = \begin{cases} ce^{-x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

For what value of c is this a valid probability density function?
What is its median?

Expectation and Variance

Definition. The **expected value** of a continuous random variable X with pdf $f(x)$ is:

$$\mu = \mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$$

Example

If X is a continuous random variable with pdf, $f(x) = \frac{c}{x^3}, x \geq 1$, then $\mathbb{E}[X] =$?

Definition. The **variance** of a continuous random variable X with pdf $f(x)$ is:

$$\sigma^2 = \text{Var}[X] = \int_{-\infty}^{\infty} (x-\mu)^2 \cdot f(x)\, dx$$

$$\begin{aligned}
Var[X] &= \int_{-\infty}^{\infty} (x-\mu)^2 \cdot f(x)\, dx \\
&= \int_{-\infty}^{\infty} (x^2 - 2\mu x + \mu^2) \cdot f(x)\, dx \\
&= \int_{-\infty}^{\infty} x^2 \cdot f(x)\, dx + \int_{-\infty}^{\infty} (-2\mu x) \cdot f(x)\, dx + \int_{-\infty}^{\infty} \mu^2 \cdot f(x)\, dx \\
&= \mathbb{E}[X^2] - 2\mu \int_{-\infty}^{\infty} x \cdot f(x)\, dx + \mu^2 \cdot 1 \\
&= \mathbb{E}[X^2] - 2\mu^2 + \mu^2 \\
&= \mathbb{E}[X^2] - (\mathbb{E}[X])^2
\end{aligned}$$

Definition. If X is a continuous random variable with pdf $f(x)$ and $g : \mathbb{R} \to \mathbb{R}$ then $g(X)$ is a random variable, and

$$\mathbb{E}[g(x)] = \int_{-\infty}^{\infty} g(x) f(x)\, dx$$

Definition. The **mode** of a continuous random variable X with pdf $f(x)$ is the value of x that maximises $f(x)$.

Example

Suppose X has the piecewise (triangular) pdf:

$$f(x) = \begin{cases} x & \text{if } 0 \leq x \leq 1 \\ 2-x & \text{if } 1 < x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

(a) Verify that $f(x)$ is a valid pdf.

(b) Find $\mathbb{P}(X \leq 0.5)$.

(c) Find $\mathbb{E}[X]$.

(d) Find $\text{Var}[X]$.

Cumulative Distribution Function

Definition. The **cumulative distribution function (CDF)** of a continuous random variable X with pdf f is defined by:

$$F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(t)\,dt$$

Note that $F'(x) = f(x)$, and $F(-\infty) = 0$, $F(\infty) = 1$.

Definition. The **median** m of X satisfies $F(m) = 0.5$.
The **lower quartile** Q_1 satisfies $F(Q_1) = 0.25$, and the **upper quartile** Q_3 satisfies $F(Q_3) = 0.75$.

Example

Suppose X has pdf $f(x) = 2x$ for $0 \leq x \leq 1$ (and 0 otherwise).

(a) Find the CDF $F(x)$.

(b) Find the median of X.

(c) Find the interquartile range of X.

Coding

If X is a random variable and $Y = aX + b$, then Y is a random variable.

Theorem

$$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$$
$$\mathrm{Var}[aX + b] = a^2\mathrm{Var}[X]$$

Continuous uniform distribution

Example

Suppose buses arrive once every 10 minutes, and we don't know the time. How long should we expect to wait for the bus?

Definition. A random variable $X \sim U(a,b)$ is distributed with the **continuous uniform distribution** on $[a,b]$ if its pdf $f(x)$ is

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{if } a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

Fact — If $X \sim U(a,b)$

$$\mu = \mathbb{E}[X] = \frac{a+b}{2}$$
$$\sigma^2 = \mathrm{Var}[X] = \frac{(b-a)^2}{12}$$

The Normal Distribution

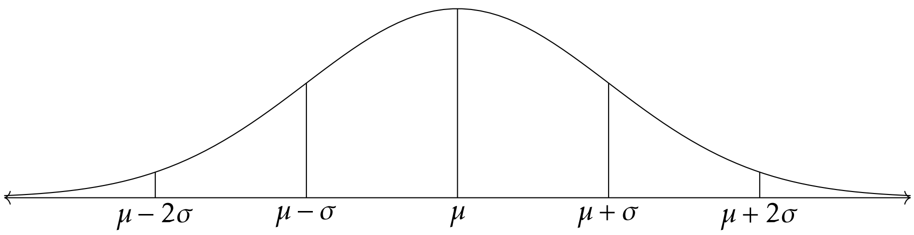
Example

Suppose we looked at the height of all the people in the country, what would it look like?

Definition. A random variable $X \sim N(\mu, \sigma^2)$ is **normally distributed** with mean μ and standard deviation σ if its pdf $f(x)$ is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

Fact — If $X \sim N(\mu, \sigma^2)$, $\quad \mathbb{E}[X] = \mu, \quad \sigma^2 = \mathrm{Var}[X]$

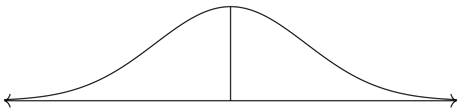


Fact (The 68–95–99.7 Rule) — For a normal distribution:

- Approximately 68% of data falls within $\mu \pm \sigma$
- Approximately 95% of data falls within $\mu \pm 2\sigma$
- Approximately 99.7% of data falls within $\mu \pm 3\sigma$

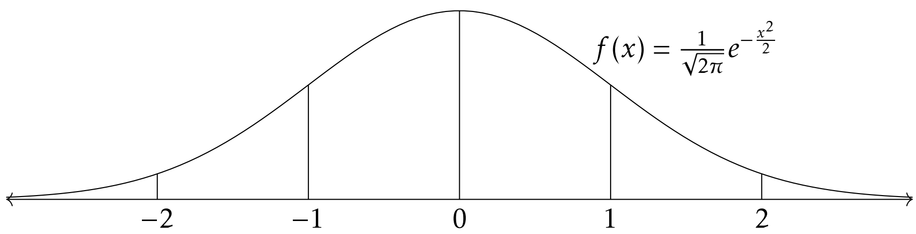
Example

Given that $X \sim N(5, 4)$, find the probability that $\mathbb{P}(2 \le X \le 6)$



Fact — If $Z \sim N(0, 1)$, we say Z is a **standard normal distribution**.

Fact — If $X \sim N(\mu, \sigma^2)$ then $\dfrac{X - \mu}{\sigma} \sim N(0, 1)$



Definition. The **standard cumulative normal distribution function** Φ is defined by $\Phi(t) = \mathbb{P}(Z \le t)$

Fact — There used to be tables of values for this. (In fact, there still are! Check your formula book)

Inverse Normal

Sometimes we need to work backwards: given a probability p, find the value z such that $\Phi(z) = p$, i.e. $\mathbb{P}(Z \le z) = p$. We write $z = \Phi^{-1}(p)$.

Example

Find the value of z such that $\mathbb{P}(Z \le z) = 0.975$.

From the tables, $\Phi(1.96) = 0.975$, so $z = 1.96$.

Example

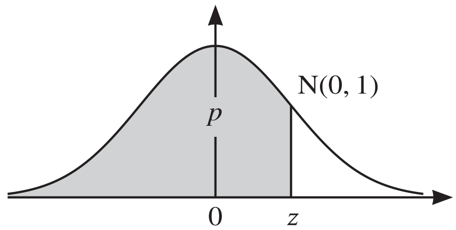
Suppose $X \sim N(50, 25)$. Find the value of x such that $\mathbb{P}(X \leq x) = 0.9$.

We need $\Phi\left(\frac{x-50}{5}\right) = 0.9$, so $\frac{x-50}{5} = \Phi^{-1}(0.9) = 1.2816$, giving $x = 56.41$ (2 d.p.).

THE NORMAL DISTRIBUTION AND ITS INVERSE

The Normal distribution: values of $\Phi(z) = p$

The table gives the probability, p, of a random variable distributed as N(0, 1) *being less than z.*



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09	1	2	3	4	5 *(add)*	6	7	8	9
0.0	.5000	5040	5080	5120	5160	5199	5239	5279	5319	5359	4	8	12	16	20	24	28	32	36
0.1	.5398	5438	5478	5517	5557	5596	5636	5675	5714	5753	4	8	12	16	20	24	28	32	35
0.2	.5793	5832	5871	5910	5948	5987	6026	6064	6103	6141	4	8	12	15	19	23	27	31	35
0.3	.6179	6217	6255	6293	6331	6368	6406	6443	6480	6517	4	8	11	15	19	23	26	30	34
0.4	.6554	6591	6628	6664	6700	6736	6772	6808	6844	6879	4	7	11	14	18	22	25	29	32
0.5	.6915	6950	6985	7019	7054	7088	7123	7157	7190	7224	3	7	10	14	17	21	24	27	31
0.6	.7257	7291	7324	7357	7389	7422	7454	7486	7517	7549	3	6	10	13	16	19	23	26	29
0.7	.7580	7611	7642	7673	7704	7734	7764	7794	7823	7852	3	6	9	12	15	18	21	24	27
0.8	.7881	7910	7939	7967	7995	8023	8051	8078	8106	8133	3	6	8	11	14	17	19	22	25
0.9	.8159	8186	8212	8238	8264	8289	8315	8340	8365	8389	3	5	8	10	13	15	18	20	23
1.0	.8413	8438	8461	8485	8508	8531	8554	8577	8599	8621	2	5	7	9	12	14	16	18	21
1.1	.8643	8665	8686	8708	8729	8749	8770	8790	8810	8830	2	4	6	8	10	12	14	16	19
1.2	.8849	8869	8888	8907	8925	8944	8962	8980	8997	9015	2	4	6	7	9	11	13	15	16
1.3	.9032	9049	9066	9082	9099	9115	9131	9147	9162	9177	2	3	5	6	8	10	11	13	14
1.4	.9192	9207	9222	9236	9251	9265	9279	9292	9306	9319	1	3	4	6	7	8	10	11	13
1.5	.9332	9345	9357	9370	9382	9394	9406	9418	9429	9441	1	2	4	5	6	7	8	10	11
1.6	.9452	9463	9474	9484	9495	9505	9515	9525	9535	9545	1	2	3	4	5	6	7	8	9
1.7	.9554	9564	9573	9582	9591	9599	9608	9616	9625	9633	1	2	3	3	4	5	6	7	8
1.8	.9641	9649	9656	9664	9671	9678	9686	9693	9699	9706	1	1	2	3	4	4	5	6	6
1.9	.9713	9719	9726	9732	9738	9744	9750	9756	9761	9767	1	1	2	2	3	4	4	5	5
2.0	.9772	9778	9783	9788	9793	9798	9803	9808	9812	9817	0	1	1	2	2	3	3	4	4
2.1	.9821	9826	9830	9834	9838	9842	9846	9850	9854	9857	0	1	1	2	2	2	3	3	4
2.2	.9861	9864	9868	9871	9875	9878	9881	9884	9887	9890	0	1	1	1	2	2	2	3	3
2.3	.9893	9896	9898	9901	9904	9906	9909	9911	9913	9916	0	1	1	1	1	2	2	2	2
2.4	.9918	9920	9922	9925	9927	9929	9931	9932	9934	9936	0	0	1	1	1	1	1	2	2
2.5	.9938	9940	9941	9943	9945	9946	9948	9949	9951	9952									
2.6	.9953	9955	9956	9957	9959	9960	9961	9962	9963	9964									
2.7	.9965	9966	9967	9968	9969	9970	9971	9972	9973	9974									
2.8	.9974	9975	9976	9977	9977	9978	9979	9979	9980	9981									
2.9	.9981	9982	9982	9983	9984	9984	9985	9985	9986	9986					*differences*				
3.0	.9987	9987	9987	9988	9988	9989	9989	9989	9990	9990					*untrustworthy*				
3.1	.9990	9991	9991	9991	9992	9992	9992	9992	9993	9993									
3.2	.9993	9993	9994	9994	9994	9994	9994	9995	9995	9995									
3.3	.9995	9995	9996	9996	9996	9996	9996	9996	9996	9997									
3.4	.9997	9997	9997	9997	9997	9997	9997	9997	9997	9998									

The Inverse Normal function: values of $\Phi^{-1}(p) = z$

p	.000	.001	.002	.003	.004	.005	.006	.007	.008	.009
.50	.0000	.0025	.0050	.0075	.0100	.0125	.0150	.0175	.0201	.0226
.51	.0251	.0276	.0301	.0326	.0351	.0376	.0401	.0426	.0451	.0476
.52	.0502	.0527	.0552	.0577	.0602	.0627	.0652	.0677	.0702	.0728
.53	.0753	.0778	.0803	.0828	.0853	.0878	.0904	.0929	.0954	.0979
.54	.1004	.1030	.1055	.1080	.1105	.1130	.1156	.1181	.1206	.1231
.55	.1257	.1282	.1307	.1332	.1358	.1383	.1408	.1434	.1459	.1484
.56	.1510	.1535	.1560	.1586	.1611	.1637	.1662	.1687	.1713	.1738
.57	.1764	.1789	.1815	.1840	.1866	.1891	.1917	.1942	.1968	.1993
.58	.2019	.2045	.2070	.2096	.2121	.2147	.2173	.2198	.2224	.2250
.59	.2275	.2301	.2327	.2353	.2378	.2404	.2430	.2456	.2482	.2508
.60	.2533	.2559	.2585	.2611	.2637	.2663	.2689	.2715	.2741	.2767
.61	.2793	.2819	.2845	.2871	.2898	.2924	.2950	.2976	.3002	.3029
.62	.3055	.3081	.3107	.3134	.3160	.3186	.3213	.3239	.3266	.3292
.63	.3319	.3345	.3372	.3398	.3425	.3451	.3478	.3505	.3531	.3558
.64	.3585	.3611	.3638	.3665	.3692	.3719	.3745	.3772	.3799	.3826
.65	.3853	.3880	.3907	.3934	.3961	.3989	.4016	.4043	.4070	.4097
.66	.4125	.4152	.4179	.4207	.4234	.4261	.4289	.4316	.4344	.4372
.67	.4399	.4427	.4454	.4482	.4510	.4538	.4565	.4593	.4621	.4649
.68	.4677	.4705	.4733	.4761	.4789	.4817	.4845	.4874	.4902	.4930
.69	.4959	.4987	.5015	.5044	.5072	.5101	.5129	.5158	.5187	.5215
.70	.5244	.5273	.5302	.5330	.5359	.5388	.5417	.5446	.5476	.5505
.71	.5534	.5563	.5592	.5622	.5651	.5681	.5710	.5740	.5769	.5799
.72	.5828	.5858	.5888	.5918	.5948	.5978	.6008	.6038	.6068	.6098
.73	.6128	.6158	.6189	.6219	.6250	.6280	.6311	.6341	.6372	.6403
.74	.6433	.6464	.6495	.6526	.6557	.6588	.6620	.6651	.6682	.6713
.75	.6745	.6776	.6808	.6840	.6871	.6903	.6935	.6967	.6999	.7031
.76	.7063	.7095	.7128	.7160	.7192	.7225	.7257	.7290	.7323	.7356
.77	.7388	.7421	.7454	.7488	.7521	.7554	.7588	.7621	.7655	.7688
.78	.7722	.7756	.7790	.7824	.7858	.7892	.7926	.7961	.7995	.8030
.79	.8064	.8099	.8134	.8169	.8204	.8239	.8274	.8310	.8345	.8381
.80	.8416	.8452	.8488	.8524	.8560	.8596	.8633	.8669	.8705	.8742
.81	.8779	.8816	.8853	.8890	.8927	.8965	.9002	.9040	.9078	.9116
.82	.9154	.9192	.9230	.9269	.9307	.9346	.9385	.9424	.9463	.9502
.83	.9542	.9581	.9621	.9661	.9701	.9741	.9782	.9822	.9863	.9904
.84	.9945	.9986	1.003	1.007	1.011	1.015	1.019	1.024	1.028	1.032
.85	1.036	1.041	1.045	1.049	1.054	1.058	1.063	1.067	1.071	1.076
.86	1.080	1.085	1.089	1.094	1.099	1.103	1.108	1.112	1.117	1.122
.87	1.126	1.131	1.136	1.141	1.146	1.150	1.155	1.160	1.165	1.170
.88	1.175	1.180	1.185	1.190	1.195	1.200	1.206	1.211	1.216	1.221
.89	1.227	1.232	1.237	1.243	1.248	1.254	1.259	1.265	1.270	1.276
.90	1.282	1.287	1.293	1.299	1.305	1.311	1.317	1.323	1.329	1.335
.91	1.341	1.347	1.353	1.360	1.366	1.372	1.379	1.385	1.392	1.398
.92	1.405	1.412	1.419	1.426	1.433	1.440	1.447	1.454	1.461	1.468
.93	1.476	1.483	1.491	1.499	1.506	1.514	1.522	1.530	1.538	1.546
.94	1.555	1.563	1.572	1.581	1.589	1.598	1.607	1.616	1.626	1.635
.95	1.645	1.655	1.665	1.675	1.685	1.695	1.706	1.717	1.728	1.739
.96	1.751	1.762	1.774	1.787	1.799	1.812	1.825	1.838	1.852	1.866
.97	1.881	1.896	1.911	1.927	1.943	1.960	1.977	1.995	2.014	2.034
.98	2.054	2.075	2.097	2.120	2.144	2.170	2.197	2.226	2.257	2.290
.99	2.326	2.366	2.409	2.457	2.512	2.576	2.652	2.748	2.878	3.090

Example

Suppose $Z \sim N(0,1)$. Find the probabilities:

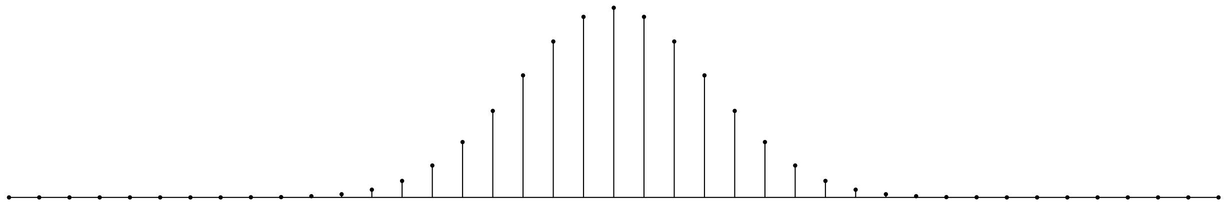
(a) $\mathbb{P}(Z < 1.41)$

(b) $\mathbb{P}(Z < -0.5)$

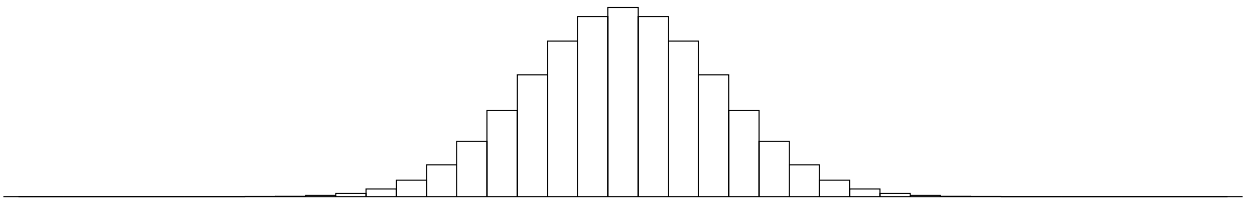
(c) $\mathbb{P}(Z > 1.975)$

Approximating Binomial to Normal distribution

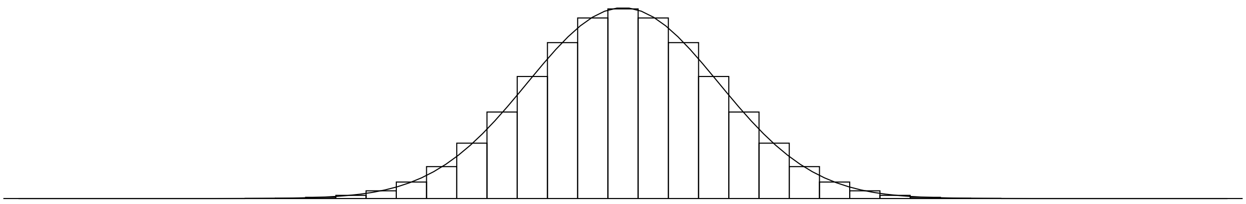
Suppose we plotted $B(40, 0.5)$, we would end up with something which looks like this



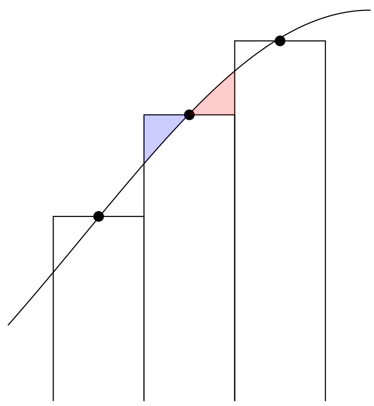
To compare this to a continuous distribution, let's replace a point mass at each integer with a uniform distribution from $n - \frac{1}{2}, n + \frac{1}{2}$



Suppose we sketched a normal distribution over the top of this distribution:



A surprisingly good match?
Let's zoom in on a specific point



Assuming the red and blue areas are equal, then it will be a perfect match.

Fact — If $X \sim B(n, p)$ is distributed with binomially, then we can make some approximations, using $Y \sim N(np, npq)$

$$\mathbb{P}(X < k) = \mathbb{P}(Y \leq k - 0.5) = \Phi\left(\frac{k - np - 0.5}{\sqrt{npq}}\right)$$

$$\mathbb{P}(X \leq k) = \mathbb{P}(Y \leq k + 0.5) = \Phi\left(\frac{k - np + 0.5}{\sqrt{npq}}\right)$$

Example (without a calculator!)

Suppose $X \sim B(170, \frac{1}{2})$, using a normal approximation, estimate the values:

- $\mathbb{P}(X \leq 91)$
- $\mathbb{P}(78 \leq X \leq 91)$

Suppose $X \sim B(121, \frac{1}{2})$, using a normal approximation, estimate the values:

- $\mathbb{P}(X < 72)$
- $\mathbb{P}(50 < X < 72)$

Tip

OCR suggests that we can use approximate a Binomial with a normal when $np > 5$ and $nq > 5$

Example (STEP 1 2013 Q12)

Each day, I have to take k different types of medicine, one tablet of each. The tablets are identical in appearance. When I go on holiday for n days, I put n tablets of each type in a container and on each day of the holiday I select k tablets at random from the container.

(i) In the case $k = 3$, show that the probability that I will select one tablet of each type on the first day of a three-day holiday is $\frac{9}{28}$. Write down the probability that I will be left with one tablet of each type on the last day (irrespective of the tablets I select on the first day).

(i) In the case $k = 3$, find the probability that I will select one tablet of each type on the first day of an n-day holiday.

(i) In the case $k = 2$, find the probability that I will select one tablet of each type on each day of an n-day holiday, and use Stirling's approximation

$$n! \approx \sqrt{2n\pi}\left(\frac{n}{e}\right)^n$$

to show that this probability is approximately $2^{-n}\sqrt{n\pi}$.

(i) The probability the first is different to the second is $\frac{6}{8}$, the probability the third is different to both of the first two is $\frac{3}{7}$ therefore the probability is $\frac{6}{8} \cdot \frac{3}{7} = \frac{9}{28}$

Whatever pills we are left with on the last day is essentially the same random choice as we make on the first day, therefore $\frac{9}{28}$

(ii) The probability the first is different to the second is $\frac{2n}{3n-1}$, the probability the third is different to both of the first two is $\frac{n}{3n-2}$ therefore the probability is $\frac{2n^2}{(3n-1)(3n-2)}$.

[We can also view this as $\frac{(3n)\cdot(2n)\cdot n}{(3n)\cdot(3n-1)\cdot(3n-2)}$]

(iii) Suppose describe the pills as B and R and also number them, then we must have a sequence of the form:

$$B_1R_1\,B_2R_2\,B_3R_3\cdots B_nR_n$$

However, we can also rearrange the order of the B and R pills in $n!$ ways each, and also the order of the pairs in 2^n ways. There are $(2n)!$ orders we could have taken the pills out therefore the probability is

$$\begin{aligned} P &= \frac{2^n(n!)^2}{(2n)!} = \frac{2^n}{\binom{2n}{n}} \\ &\approx \frac{2^n \cdot 2n\pi\left(\frac{n}{e}\right)^{2n}}{\sqrt{2\cdot 2n\cdot\pi}\left(\frac{2n}{e}\right)^{2n}} \\ &= \frac{2^n\sqrt{n\pi}\cdot n^{2n}\cdot e^{-2n}}{2^{2n}\cdot n^{2n}\cdot e^{-2n}} \\ &= 2^{-n}\sqrt{n\pi} \end{aligned}$$

There is a nice way to think about this question using conditional probability. Suppose we are drawing out of an infinitely supply of R and B pills, then each day there is a $\frac{1}{2}$ chance of getting different pills. Therefore over n days there is a 2^{-n} chance of getting different pills. Conditional on the balanced total we see that

$$\mathbb{P}(\textit{balanced every day}|\textit{balanced total}) = \frac{\mathbb{P}(\textit{balanced every day})}{\mathbb{P}(\textit{balanced total})}$$

We have already seen the term that is balanced total is $\frac{1}{2^{2n}}\binom{2n}{n}$*, but we can also approximate the balanced total using a normal approximation.* $B(2n, \frac{1}{2}) \approx N(n, \frac{n}{2})$ *and so:*

$$\begin{aligned}
\mathbb{P}(X = n) &\approx \mathbb{P}\left(n - 0.5 \leq \sqrt{\tfrac{n}{2}}Z + n \leq n + 0.5\right) \\
&= \mathbb{P}\left(-\frac{1}{\sqrt{2n}} \leq Z \leq \frac{1}{\sqrt{2n}}\right) \\
&= \int_{-\frac{1}{\sqrt{2n}}}^{\frac{1}{\sqrt{2n}}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \mathrm{d}x \approx \frac{2}{\sqrt{2n}} \frac{1}{\sqrt{2\pi}} \\
&\approx \frac{1}{\sqrt{n\pi}}
\end{aligned}$$