1 Predictable Gaussians

Note 20

Let Y be the result of a fair coin flip, and X be a normally distributed random variable with parameters dependent on Y. That is, if $Y = 1$, then $X \sim N(\mu_1, \sigma_1^2)$, and if $Y = 0$, then $X \sim N(\mu_0, \sigma_0^2)$.

(a) Sketch the two distributions of X overlaid on the same graph for the following cases:

(i) $\sigma_0^2 = \sigma_1^2, \mu_0 \neq \mu_1$

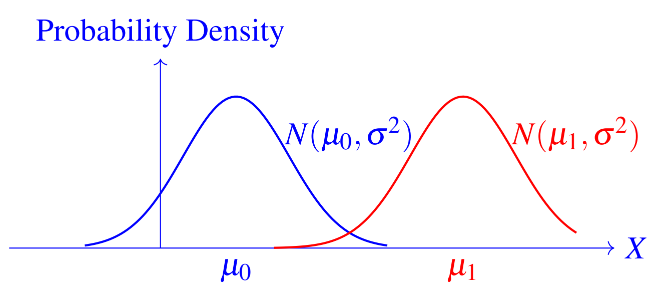
(ii) $\sigma_0^2 \neq \sigma_1^2, \mu_0 = \mu_1$

(b) Bayes' rule for mixed distributions can be formulated as $\mathbb{P}[Y = 1 \mid X = x] = \frac{\mathbb{P}[Y=1] f_{X|Y=1}(x)}{f_X(x)}$ where Y is a discrete distribution and X is a continuous distribution. Compute $\mathbb{P}[Y = 1 \mid X = x]$, and show that this can be expressed in the form of $\frac{1}{1+e^\gamma}$ for some expression γ. (Hint: any value z can be equivalently expressed as $e^{\ln(z)}$)

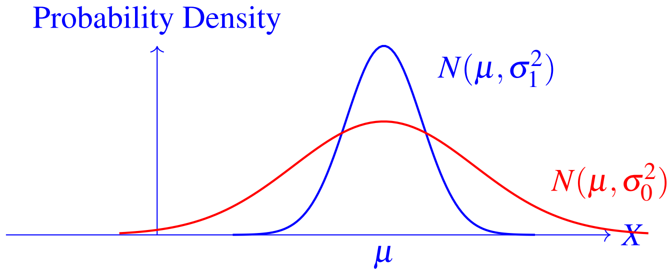
(c) In the special case where $\sigma_0^2 = \sigma_1^2$ find a simple expression for the value of x where $\mathbb{P}[Y = 1 \mid X = x] = \mathbb{P}[Y = 0 \mid X = x] = 1/2$, and interpret what the expression represents. (Hint: the identity $(a+b)(a-b) = a^2 - b^2$ may be useful)

Solution:

(a) (i) In this case, there are two bell curves with the same spread/width due to the variances being equal, but being centered at different means.



(ii) In this case, there will be two bell curves centered at the same mean, but the one with lower variance will be skinnier and taller, due to more of the probability density being centered closer to the mean.



(b)

$$
\begin{aligned}
&\mathbb{P}[Y = 1 \mid X = x] \\
&= \frac{\mathbb{P}[Y=1] f_{X|Y=1}(x)}{\mathbb{P}[Y=1] f_{X|Y=1}(x) + \mathbb{P}[Y=0] f_{X|Y=0}(x)} \\
&= \frac{\frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right)}{\frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right) + \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{(x-\mu_0)^2}{2\sigma_0^2}\right)} \\
&= \frac{1}{1 + \frac{\sigma_1}{\sigma_0} \exp\left(\frac{(x-\mu_1)^2}{2\sigma_1^2} - \frac{(x-\mu_0)^2}{2\sigma_0^2}\right)} \\
&= \frac{1}{1 + \exp\left(\ln\left(\frac{\sigma_1}{\sigma_0}\right) + \frac{(x-\mu_1)^2}{2\sigma_1^2} - \frac{(x-\mu_0)^2}{2\sigma_0^2}\right)}.
\end{aligned}
$$

Which is of the desired form, with $\gamma = \ln\left(\frac{\sigma_1}{\sigma_0}\right) + \left(\frac{(x-\mu_1)^2}{2\sigma_1^2} - \frac{(x-\mu_0)^2}{2\sigma_0^2}\right)$

(c) Note that $\mathbb{P}[Y = 1 \mid X = x] = \frac{1}{2}$ implies that $\exp(\gamma) = 1$, which means that $\gamma = 0$. Thus, $\ln\left(\frac{\sigma_1}{\sigma_0}\right) + \left(\frac{(x-\mu_1)^2}{2\sigma_1^2} - \frac{(x-\mu_0)^2}{2\sigma_0^2}\right) = 0$. Using the conditions from the problem statement, we can simplify this expression.

$$
\begin{aligned}
&\ln\left(\frac{\sigma_1}{\sigma_0}\right) + \left(\frac{(x-\mu_1)^2}{2\sigma^2} - \frac{(x-\mu_0)^2}{2\sigma^2}\right) = 0 \\
&0 + \left(\frac{(x-\mu_1)^2}{2\sigma^2} - \frac{(x-\mu_0)^2}{2\sigma^2}\right) = 0 \\
&(x-\mu_1)^2 = (x-\mu_0)^2 \\
&x^2 - 2\mu_1 x + \mu_1^2 = x^2 - 2\mu_0 x + \mu_0^2 \\
&2(\mu_0 - \mu_1)x = \mu_0^2 - \mu_1^2 \\
&x = \frac{\mu_0^2 - \mu_1^2}{2(\mu_0 - \mu_1)} = \frac{\mu_0 + \mu_1}{2}
\end{aligned}
$$

Notice that x becomes the average, or center, of the two means.

2 Chebyshev's Inequality vs. Central Limit Theorem

Note 19
Note 20

Let n be a positive integer. Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with the following distribution:

$$\mathbb{P}[X_i = -1] = \frac{1}{12}; \qquad \mathbb{P}[X_i = 1] = \frac{9}{12}; \qquad \mathbb{P}[X_i = 2] = \frac{2}{12}.$$

(a) Calculate the expectations and variances of X_1, $\sum_{i=1}^n X_i$, $\sum_{i=1}^n (X_i - \mathbb{E}[X_i])$, and

$$Z_n = \frac{\sum_{i=1}^n (X_i - \mathbb{E}[X_i])}{\sqrt{n/2}}.$$

(b) Use Chebyshev's Inequality to find an upper bound b for $\mathbb{P}[|Z_n| \geq 2]$.

(c) Use b from the previous part to bound $\mathbb{P}[Z_n \geq 2]$ and $\mathbb{P}[Z_n \leq -2]$.

(d) As $n \to \infty$, what is the distribution of Z_n?

(e) We know that if $Z \sim \mathcal{N}(0,1)$, then $\mathbb{P}[|Z| \leq 2] = \Phi(2) - \Phi(-2) \approx 0.9545$. As $n \to \infty$, provide approximations for $\mathbb{P}[Z_n \geq 2]$ and $\mathbb{P}[Z_n \leq -2]$.

Solution:

(a) Firstly, let us calculate $\mathbb{E}[X_1]$ and $\text{Var}(X_1)$; we have

$$\mathbb{E}[X_1] = -\frac{1}{12} + \frac{9}{12} + \frac{4}{12} = 1$$
$$\text{Var}(X_1) = \frac{1}{12} \cdot 2^2 + \frac{9}{12} \cdot 0^2 + \frac{2}{12} \cdot 1^2 = \frac{1}{2}.$$

Using linearity of expectation and variance (since $X_1, \ldots, X_n$ are independent), we find that

$$\mathbb{E}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbb{E}[X_i] = n$$
$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) = \frac{n}{2}$$

Again, by linearity of expectation,

$$\mathbb{E}\left[\sum_{i=1}^n (X_i - \mathbb{E}[X_i])\right] = \mathbb{E}\left[\sum_{i=1}^n X_i - n\right] = n - n = 0.$$

Subtracting a constant does not change the variance, so

$$\text{Var}\left(\sum_{i=1}^n (X_i - \mathbb{E}[X_i])\right) = \text{Var}\left(\sum_{i=1}^n X_i - n\right) = \frac{n}{2},$$

as before.

Using the scaling properties of the expectation and variance, we finally have

$$\mathbb{E}[Z_n] = \mathbb{E}\left[\frac{\sum_{i=1}^n (X_i - \mathbb{E}[X_i])}{\sqrt{n/2}}\right] = \frac{1}{\sqrt{n/2}}\mathbb{E}\left[\sum_{i=1}^n (X_i - \mathbb{E}[X_i])\right] = \frac{0}{\sqrt{n/2}} = 0$$
$$\mathrm{Var}(Z_n) = \mathrm{Var}\left(\frac{\sum_{i=1}^n (X_i - \mathbb{E}[X_i])}{\sqrt{n/2}}\right) = \frac{1}{n/2}\mathrm{Var}\left(\sum_{i=1}^n (X_i - \mathbb{E}[X_i])\right) = \frac{n/2}{n/2} = 1$$

(b) Using Chebyshev's, we have

$$\mathbb{P}[|Z_n| \geq 2] \leq \frac{\mathrm{Var}(Z_n)}{2^2} = \frac{1}{4}$$

since $\mathbb{E}[Z_n] = 0$ and $\mathrm{Var}(Z_n) = 1$ as we computed in the previous part.

(c) $\frac{1}{4}$ for both, since we have

$$\mathbb{P}[Z_n \geq 2] \leq \mathbb{P}[|Z_n| \geq 2]$$
$$\mathbb{P}[Z_n \leq -2] \leq \mathbb{P}[|Z_n| \geq 2]$$

(d) By the Central Limit Theorem, we know that $Z_n \to \mathcal{N}(0,1)$, the standard normal distribution.

(e) Since $Z_n \to \mathcal{N}(0,1)$, we can approximate $\mathbb{P}[|Z_n| \geq 2] \approx 1 - 0.9545 = 0.0455$. By the symmetry of the normal distribution, $\mathbb{P}[Z_n \geq 2] = \mathbb{P}[Z_n \leq -2] \approx 0.0455/2 = 0.02275$.

It is interesting to note that the CLT provides a much smaller answer than Chebyshev. This is due to the fact that the CLT is applied to a particular kind of random variable, namely the (scaled) sum of a bunch of i.i.d. random variables. Chebyshev's inequality, however, holds for any random variable, and is therefore weaker.

3 Law of Large Numbers

Recall that the *Law of Large Numbers* holds if, for every $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left[\left|\frac{1}{n}S_n - \mathbb{E}\left[\frac{1}{n}S_n\right]\right| > \varepsilon\right] = 0.$$

In class, we saw that the Law of Large Numbers holds for $S_n = X_1 + \cdots + X_n$, where the X_i's are i.i.d. random variables. This problem explores if the Law of Large Numbers holds under other circumstances.

Packets are sent from a source to a destination node over the Internet. Each packet is sent on a certain route, and the routes are disjoint. Each route has a failure probability of $p \in (0,1)$ and different routes fail independently. If a route fails, all packets sent along that route are lost. You can assume that the routing protocol has no knowledge of which route fails.

For each of the following routing protocols, determine whether the Law of Large Numbers holds when S_n is defined as the total number of received packets out of n packets sent. Answer **Yes** if the Law of Large Number holds, or **No** if not. Give a justification of your answer. (Whenever convenient, you can assume that n is even.)

(a) **Yes** or **No**: Each packet is sent on a completely different route.

(b) **Yes** or **No**: The packets are split into $n/2$ pairs of packets. Each pair is sent together on its own route (i.e., different pairs are sent on different routes).

(c) **Yes** or **No**: The packets are split into 2 groups of $n/2$ packets. All the packets in each group are sent on the same route, and the two groups are sent on different routes.

(d) **Yes** or **No**: All the packets are sent on one route.

Solution:

(a) **Yes.** Define X_i to be 1 if a packet is sent successfully on route i. Then X_i, $i = 1, \ldots, n$ is 0 with probability p and 1 otherwise. Since we have individual routes for each packet, we have a total of n routes. The total number of successful packets sent is hence $S_n = X_1 + \cdots + X_n$. Since S_n is a sum of i.i.d. Bernoulli random variables, $S_n \sim \text{Binomial}(n, 1-p)$.

Now similar to notation in the lecture notes, we define $A_n = S_n/n$ to be the fraction of successful packets sent, out of the n packets. Moreover, for each X_i,

$$\mathbb{E}[X_i] = 1 - p$$

and

$$\text{Var}(X_i) = p(1-p).$$

Using Chebyshev's inequality:

$$\mathbb{P}[|A_n - \mathbb{E}[A_n]| > \varepsilon] = \mathbb{P}[|A_n - (1-p)| > \varepsilon] \le \frac{\text{Var}[A_n]}{\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2} \to 0 \quad \text{as } n \to \infty.$$

(b) **Yes.** Now we need $n/2$ routes for each pair of packets. Similarly to the previous question, we define X_i, $i = 1, \ldots, n/2$ to be 0 with probability p and 2 (packets) otherwise. Now the total number of packets is $S_n = X_1 + \cdots + X_{n/2}$ and the fraction of received packets is $A_n = S_n/n$.

Now for each $i = 1, \ldots, n/2$,

$$\mathbb{E}[X_i] = 2(1-p)$$

and

$$\text{Var}(X_i) = 4p(1-p).$$

Thus,

$$\mathbb{E}[A_n] = \frac{\mathbb{E}[X_1] + \ldots + \mathbb{E}[X_{n/2}]}{n} = \frac{1}{n} \cdot \frac{n}{2} \cdot 2(1-p) = 1 - p$$

and

$$\text{Var}(A_n) = \frac{1}{n^2}\big(\text{Var}(X_1) + \ldots + \text{Var}(X_{n/2})\big) = \frac{1}{n^2} \cdot \frac{n}{2} 4p(1-p) = \frac{2p(1-p)}{n}.$$

Finally, we get:

$$\mathbb{P}[|A_n - \mathbb{E}[A_n]| > \varepsilon] = \mathbb{P}[|A_n - (1-p)| > \varepsilon] \le \frac{2p(1-p)}{n\varepsilon^2} \to 0 \quad \text{as } n \to \infty.$$

(c) **No.** In this situation, we have that no packets get through with probability p^2, half the packets get through with probability $2p(1-p)$, and all the packets get through with probability $(1-p)^2$. This tells us that $\frac{1}{n}S_n$ is 0 with probability p^2, $\frac{1}{2}$ with probability $2p(1-p)$, and 1 with probability $(1-p)^2$. Since $\mathbb{E}\left[\frac{1}{n}S_n\right] = 1-p$, this gives us that

$$\left|\frac{1}{n}S_n - \mathbb{E}\left[\frac{1}{n}S_n\right]\right| = \begin{cases} 1-p & \text{with probability } p^2 \\ \left|p-\frac{1}{2}\right| & \text{with probability } 2p(1-p) \\ p & \text{with probability } (1-p)^2 \end{cases}$$

We now consider two cases: either $p = \frac{1}{2}$ or $p \neq \frac{1}{2}$. In the former case, we can take $\varepsilon = \frac{1}{4}$, and we'll have that

$$\begin{aligned} \mathbb{P}\left[\left|\frac{1}{n}S_n - \mathbb{E}\left[\frac{1}{n}S_n\right]\right| > \varepsilon\right] &= \mathbb{P}\left[\frac{1}{n}S_n = 0 \cup \frac{1}{n}S_n = 1\right] \\ &= \frac{1}{2} \end{aligned}$$

In the latter case, we can take $\varepsilon = \frac{\min\left(1-p, \left|p-\frac{1}{2}\right|, p\right)}{2}$ and we'll have that

$$\mathbb{P}\left[\left|\frac{1}{n}S_n - \mathbb{E}\left[\frac{1}{n}S_n\right]\right| > \varepsilon\right] = 1$$

Since neither of these probabilities converge to zero as $n \to \infty$, we have that the WLLN does not hold in either case.

(d) **No.** In this case, we have that no packets get through with probability p and all the packets get through with probability $(1-p)$. Hence,

$$\left|\frac{1}{n}S_n - \mathbb{E}\left[\frac{1}{n}S_n\right]\right| = \begin{cases} 1-p & \text{with probability } p \\ p & \text{with probability } (1-p) \end{cases}$$

So if we take $\varepsilon = \frac{\min(p, 1-p)}{2}$, we have that

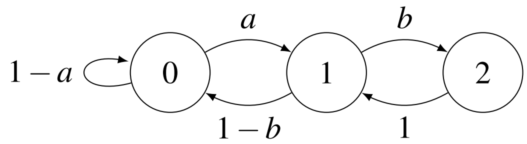
$$\mathbb{P}\left[\left|\frac{1}{n}S_n - \mathbb{E}\left[\frac{1}{n}S_n\right]\right| > \varepsilon\right] = 1$$

As in the previous part, because this does not converge to 0 as $n \to \infty$, we have that the WLLN does not hold.

For problems (c) and (d), you should've had the intuition that since the packets are automatically sent through 1 or 2 routes, increasing n does not really help for LLN.

4 Analyze a Markov Chain

Note 21 Consider a Markov chain with the state diagram shown below where $a, b \in (0, 1)$.



Here, we let $X(n)$ denote the state at time n.

(a) Is this Markov chain irreducible? Is this Markov chain aperiodic? Justify your answers.

(b) Calculate $\mathbb{P}[X(1)=1, X(2)=0, X(3)=1, X(4)=2 \mid X(0)=0]$.

(c) Calculate the invariant distribution. Do all initial distributions converge to this invariant distribution? Justify your answer.

Solution:

(a) The Markov chain is irreducible because $a, b \in (0,1)$. Also, $P(0,0) > 0$, so that

$$\gcd\{n > 0 \mid P^n(0,0) > 0\} = \gcd\{1,2,3,\ldots\} = 1,$$

which shows that the Markov chain is aperiodic.

We can also notice from the definition of aperiodicity that if a Markov chain has a self loop with nonzero probability, it is aperiodic. In particular, a self loop implies that the smallest number of steps we need to take to get from a state back to itself is 1. In this case, since $P(0,0) > 0$, we have a self loop with nonzero probability, which makes the Markov chain aperiodic.

(b) As a result of the Markov property, we know our state at timestep n depends only on timestep $n-1$. Looking at the transition probabilities, we see that the final expression is

$$P(0,1) \times P(1,0) \times P(0,1) \times P(1,2) = a(1-b)ab.$$

(c) The balance equations are

$$\begin{cases} \pi(0) = (1-a)\pi(0) + (1-b)\pi(1) \\ \pi(1) = a\pi(0) + \pi(2) \end{cases} \implies \begin{cases} a\pi(0) = (1-b)\pi(1) \\ \pi(1) = a\pi(0) + \pi(2) \end{cases}$$

$$\implies \begin{cases} a\pi(0) = (1-b)\pi(1) \\ \pi(1) = a\left(\frac{1-b}{a}\pi(1)\right) + \pi(2) \end{cases}$$

$$\implies \begin{cases} a\pi(0) = (1-b)\pi(1) \\ b\pi(1) = \pi(2) \end{cases}$$

As a side note, these last equations express the equality of the probability of a jump from i to $i+1$ and from $i+1$ to i, for $i=0$ and $i=1$, respectively. These relations are also called the "detailed balance equations".

From these equations we find successively that

$$\pi(1) = \frac{a}{1-b}\pi(0) \qquad \pi(2) = b\pi(1) = \frac{ab}{1-b}\pi(0).$$

The normalization equation is

$$1 = \pi(0) + \pi(1) + \pi(2) = \pi(0)\left(1 + \frac{a}{1-b} + \frac{ab}{1-b}\right)$$

$$1 = \pi(0)\left(\frac{1-b+a+ab}{1-b}\right)$$

so that

$$\pi(0) = \frac{1-b}{1-b+a+ab}.$$

Thus,

$$\pi(0) = \frac{1-b}{1-b+a+ab} \qquad \pi(1) = \frac{a}{1-b+a+ab} \qquad \pi(2) = \frac{ab}{1-b+a+ab}$$

Or in vector form,

$$\pi = \frac{1}{1-b+a+ab}\begin{bmatrix}1-b & a & ab\end{bmatrix}.$$

Since the Markov chain is irreducible and aperiodic, all initial distributions converge to this invariant distribution by the fundamental theorem of Markov chains.

5 A Bit of Everything

Note 21

Suppose that $X_0, X_1, \ldots$ is a Markov chain with finite state space $S = \{1, 2, \ldots, n\}$, where $n > 2$, and transition matrix P. Suppose further that

$$P(1,i) = \frac{1}{n} \quad \text{for all states } i \text{ and}$$
$$P(j, j-1) = 1 \quad \text{for all states } j \neq 1,$$

with $P(i,j) = 0$ everywhere else.

(a) Prove that this Markov chain is irreducible and aperiodic.

(b) Suppose you start at state 1. What is the distribution of T, where T is the number of transitions until you leave state 1 for the first time?

(c) Again starting from state 1, what is the expected number of transitions until you reach state n for the first time?

(d) Again starting from state 1, what is the probability you reach state 2 before you reach state n?

(e) Compute the stationary distribution of this Markov chain.

Solution:

(a) For any two states i and j, we can consider the path $(i, i-1, \ldots, 2, 1, j)$, which has nonzero probability of occurring. Thus, this chain is irreducible. To see that it is aperiodic, observe that $d(1) = 1$, as we have self-loop from state 1 to itself.

(b) At any given transition, we leave state 1 with probability with probability $\frac{n-1}{n}$, independently of any previous transition. Thus, the distribution is Geometric, with parameter $\frac{n-1}{n}$.

(c) Suppose that $\beta(i)$ is the expected number of transitions necessary to reach state n for the first time, starting from state i. We have the following first step equations:

$$\beta(1) = 1 + \sum_{j=1}^{n} \frac{1}{n}\beta(j),$$
$$\beta(i) = 1 + \beta(i-1) \quad \text{for } 1 < i < n \text{, and}$$
$$\beta(n) = 0.$$

We can simplify the second recurrence to

$$\beta(i) = i - 1 + \beta(1) \quad \text{for } 1 < i < n.$$

Substituting this simplified recurrence into the first equation, we get that

$$\beta(1) = 1 + \frac{1}{n}\sum_{i=1}^{n-1}(i-1+\beta(1)) = 1 + \frac{1}{n}\sum_{i=1}^{n-1}(i-1) + \frac{1}{n}\sum_{i=1}^{n-1}\beta(1) = 1 + \frac{(n-2)(n-1)}{2n} + \frac{n-1}{n}\beta(1),$$

which we can solve to get that

$$\beta(1) = \boxed{n + \frac{1}{2}(n-1)(n-2)}.$$

(d) Suppose that $\alpha(i)$ is the probability that we reach state 2 before we reach state n, starting from state i. One immediate observation we can make is that from any state i in $\{2,\ldots,n-1\}$, we are guaranteed to see state 2 before state n, as we can only take the path $(i, i-1, \ldots, 2, 1)$. Hence, $\alpha(i) = 1$ if $i \in \{2,\ldots,n-1\}$. Moreover, $\alpha(n) = 0$, so

$$\alpha(1) = \sum_{i=1}^{n} \frac{1}{n}\alpha(i) = \frac{1}{n}\alpha(1) + \sum_{i=2}^{n-1}\frac{1}{n}1 + \frac{1}{n}0 = \frac{1}{n}\alpha(1) + \frac{1}{n}(n-2),$$

hence $\alpha(1) = \boxed{\frac{n-2}{n-1}}$.

(e) We have the balance equations

$$\pi(i) = \frac{1}{n}\pi(1) + \pi(i+1) \quad \text{if } i \neq n \text{, and}$$
$$\pi(n) = \frac{1}{n}\pi(1).$$

We can collapse the first recurrence to

$$\pi(i) = \frac{n-i}{n}\pi(1) + \pi(n) = \frac{n-i+1}{n}\pi(1),$$

so we can express each stationary probability in terms of the stationary probability of state 1. We can finish by using the normalization equation:

$$\pi(1) + \pi(2) + \cdots + \pi(n) = 1 \implies \frac{1}{n}\pi(1)\sum_{i=1}^{n} n - i + 1 = 1.$$

The last sum can be rearranged to be the sum of the integers from 1 up to n, so we get that

$$\pi(1) = \frac{2}{n+1} \implies \pi = \boxed{\frac{2}{n(n+1)}\begin{bmatrix} n & n-1 & \cdots & 1 \end{bmatrix}}.$$