Avd. Matematisk statistik

KTH Matematik

TENTAMEN I SF2940 SANNOLIKHETSTEORI/EXAM IN SF2940 PROBABILITY THEORY WEDNESDAY THE 13^{th} OF JANUARY 2010 2.00 p.m.–7.00 p.m.

Examinator: Timo Koski, tel. 790 71 34, email: tjtkoski@kth.se

Tillåtna hjälpmedel Means of assistance permitted: Appendix 2 in A.Gut: An Intermediate Course in Probability. Formulas for probability theory SF2940. Pocket calculator.

You should define and explain your notation. Your computations and your line of reasoning should be written down so that they are easy to follow. Numerical values should be given with the precision of two decimal points. You may apply results stated in a part of an exam question to another part of the exam question even if you have not solved the first part. The number of exam questions (Uppgift) is six(6).

Each question gives maximum ten (10) points. 30 points will guarantee a passing result. The grade Fx (the exam can completed by extra examination) for those with 27−29 points.

Solutions to the exam questions will be available at
`http://www.math.kth.se/matstat/gru/sf2940/`
starting from Wednesday 13^{th} of January 2010 at 7.05 p.m..

The exam results will be announced at the latest on Friday the 22^{nd} of January 2010.

Your exam paper will be retainable at elevexpeditionen during a period of seven weeks after the date of the exam.

LYCKA TILL!

- - - - - - - - - - - - - - - - - - - -

The following formulas may turn out to be useful in the sequel.
Let $\mathbf{X} = (X_1, X_2, \ldots, X_m)^T$ have the probability density $f_{\mathbf{X}}(x_1, x_2, \ldots, x_m)$. Define a new random vector $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_m)^T$ by

$$Y_i = g_i(X_1, \ldots, X_m), \quad i = 1, 2, \ldots, m,$$

where g_i are continuously differentiable and $(g_1, g_2, \ldots, g_m)$ is invertible (in a domain) with

$$X_i = h_i(Y_1, \ldots, Y_m), \quad i = 1, 2, \ldots, m,$$

where h_i are continuously differentiable. Then the density of $\mathbf{Y}$ is (in the domain of invertibility)

$$f_{\mathbf{Y}}(y_1, \ldots, y_m) = f_{\mathbf{X}}(h_1(y_1, y_2, \ldots, y_m), \ldots, h_m(y_1, y_2, \ldots, y_m)) \mid J \mid,$$

where J is the Jacobian determinant

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_m} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_m} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial x_m}{\partial y_1} & \frac{\partial x_m}{\partial y_2} & \cdots & \frac{\partial x_m}{\partial y_m} \end{vmatrix}.$$

Uppgift 1

The column vector $(X_1, X_2)^T$ has a bivariate normal distribution $N_2(\mathbf{0}, \Sigma)$, where $\mathbf{0} = (0,0)^T$, i.e.,

$$f_{\mathbf{X}}(x_1, x_2) = \frac{1}{2\pi\sqrt{\det \Sigma}} e^{-\frac{1}{2}\mathbf{x}^T \Sigma^{-1} \mathbf{x}},$$

where $\mathbf{x}^T = (x_1, x_2)$. We introduce two new random variables by

$$\begin{pmatrix} Y_1 \\ \\ Y_2 \end{pmatrix} = \begin{pmatrix} \frac{e^{X_1}}{1+e^{X_1}+e^{X_2}} \\ \\ \frac{e^{X_2}}{1+e^{X_1}+e^{X_2}} \end{pmatrix}.$$

Find the probability density of $(Y_1, Y_2)^T$. (10 p)

Aid: It is maybe useful to observe that

$$1 - (Y_1 + Y_2) = \frac{1}{1 + e^{X_1} + e^{X_2}}.$$

Uppgift 2

Let the random variables Z_i, $i = 1, 2, \ldots$ be independent and identically Laplace distributed, $Z_i \in L(a)$, $a > 0$. We consider N of these random variables $Z_1, Z_2, \ldots, Z_N$, where N is a first success random variable, i.e., it has the probability function $(0 < p < 1)$

$$P(N = z) = p_N(z) = p \cdot (1-p)^{z-1}, \quad z = 1, 2, \ldots$$

N is independent of all Z_i:s. We set

$$S_N = Z_1 + Z_2 + \ldots + Z_N,$$

and

$$Y = c \cdot S_N.$$

(a) Find the probability generating function of N. (2 p)

(b) Find the value of c that makes $Y \in L(a)$. (8 p)

Hint: Compute the characteristic function of Y.

Uppgift 3

$W = \{W(t) \mid 0 \leq t < \infty\}$ is a Wiener process. We set for $\lambda > 0$

$$X_n = \sum_{k=1}^{n} e^{-\lambda \frac{k-1}{n}} \left(W\left(\frac{k}{n}\right) - W\left(\frac{(k-1)}{n}\right) \right), \quad n \geq 1.$$

(a) Determine the distribution of X_n. Please explain the steps of your solution in detail. (5 p)

(b) Show that there is the convergence in distribution

$$X_n \xrightarrow{d} N\left(0, \int_0^1 e^{-2\lambda t} dt\right) \quad \text{as } n \to \infty.$$

Please present the steps of your solution in detail. (5 p)

Uppgift 4

Consider the sequence U_i, $i = 1, 2, \ldots,$ of independent Exp(1) -distributed random variables. Set $X_n \stackrel{d}{=} U_1 + U_2 + \ldots + U_n$. Show that

$$\lim_{n \to \infty} P\left(X_n \leq n\right) = \frac{1}{2}.$$

(10 p)

Uppgift 5

$N = \{N(t) \mid t \geq 0\}$ is a Poisson process with intensity $\lambda > 0$.

(a) Show that there is the convergence in quadratic mean

$$\frac{N(n)}{n} \xrightarrow{q} \lambda, \quad \text{as } n \to \infty. \tag{1}$$

(6 p)

(b) Explain how the limit in (1) implies the following limit:

$$e^{n\lambda\left(e^{\frac{it}{n}}-1\right)} \to e^{i\lambda t}, \quad \text{as } n \to \infty.$$

(3 p)

(c) How is the preceding related to the law of large numbers ? (1 p)

Uppgift 6

Let $+_2$ designate binary addition modulo two, i.e.,

$$0 +_2 0 = 0, 1 +_2 0 = 1, 0 +_2 1 = 1, 1 +_2 1 = 0,$$

and let $\{X_n\}_{n\geq 1}$ be a sequence of independent Be(p)- distributed random variables. $Y_0 \in$ Be (p) is independent of each and everyone of X_n. Consider the stochastic process $\{Y_n\}_{n\geq 1}$ given by

$$Y_n = Y_{n-1} +_2 X_n, n = 1, 2, \ldots, .$$

Show that

$$P\left(Y_n = 1 \mid Y_0 = 1\right) = P\left(W = \text{ even integer } \in \{0, 1, \ldots, n\}\ \right),$$

where $W \in \text{Bin}(n, p)$ (even integer= jämnt heltal in Swedish). (10 p)

Hint: Iteration of $Y_n = Y_{n-1} +_2 X_n = Y_{n-2} +_2 X_{n-1} +_2 X_n$ and so on down to Y_0, might be useful.

KTH Matematik

Avd. Matematisk statistik

SOLUTIONS TO THE EXAM WEDNESDAY THE 13^{th} OF JANUARY 2010 02.00 p.m.–07.00 p.m..

Uppgift 1

In force of

$$\begin{pmatrix} Y_1 \\ \\ Y_2 \end{pmatrix} = \begin{pmatrix} g_1(X_1, X_2) \\ \\ g_2(X_1, X_2) \end{pmatrix} = \begin{pmatrix} \frac{e^{X_1}}{1+e^{X_1}+e^{X_2}} \\ \\ \frac{e^{X_2}}{1+e^{X_1}+e^{X_2}} \end{pmatrix}$$

we get that

$$X_1 = h_1(Y_1, Y_2) = \ln Y_1 + \ln\left(1 + e^{X_1} + e^{X_2}\right) = \ln Y_1 - \ln\left(1 - (Y_1 + Y_2)\right)$$

and similarly

$$X_2 = h_2(Y_1, Y_2) = \ln Y_2 - \ln\left(1 - (Y_1 + Y_2)\right).$$

Then we find the Jacobian, or

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}.$$

Entry by entry we get

$$\frac{\partial x_1}{\partial y_1} = \frac{1}{y_1} + \frac{1}{1 - (y_1 + y_2)}$$

$$\frac{\partial x_1}{\partial y_2} = \frac{1}{1 - (y_1 + y_2)}$$

$$\frac{\partial x_2}{\partial y_1} = \frac{1}{1 - (y_1 + y_2)}$$

$$\frac{\partial x_2}{\partial y_2} = \frac{1}{y_2} + \frac{1}{1 - (y_1 + y_2)}$$

Thus, the determinant is

$$J = \frac{\partial x_1}{\partial y_1} \cdot \frac{\partial x_2}{\partial y_2} - \frac{\partial x_1}{\partial y_2} \cdot \frac{\partial x_2}{\partial y_1}$$

$$= \frac{1}{Y_1}\left(\frac{1}{y_2} + \frac{1}{1 - (y_1 + y_2)}\right) + \frac{1}{1 - (y_1 + y_2)}\left(\frac{1}{y_2} + \frac{1}{1 - (y_1 + y_2)}\right)$$

$$- \left(\frac{1}{1 - (y_1 + y_2)}\right)^2$$

$$= \frac{1}{y_1}\frac{1}{y_2} + \frac{1}{y_1}\frac{1}{1 - (y_1 + y_2)} + \frac{1}{y_2}\frac{1}{1 - (y_1 + y_2)} + \left(\frac{1}{1 - (y_1 + y_2)}\right)^2 - \left(\frac{1}{1 - (y_1 + y_2)}\right)^2$$

$$
\begin{aligned}
&= \frac{1}{y_1}\frac{1}{y_2} + \frac{1}{y_1}\frac{1}{1-(y_1+y_2)} + \frac{1}{y_2}\frac{1}{1-(y_1+y_2)} \\
&= \frac{1}{y_1}\frac{1}{y_2} + \frac{1}{1-(y_1+y_2)}\left(\frac{1}{y_1} + \frac{1}{y_2}\right) \\
&= \frac{1}{y_1}\frac{1}{y_2} + \frac{1}{1-(y_1+y_2)}\left(\frac{y_1+y_2}{y_1 y_2}\right) \\
&= \frac{1-(y_1+y_2)+y_1+y_2}{y_1 y_2\left(1-(y_1+y_2)\right)} \\
&= \frac{1}{y_1 y_2\left(1-(y_1+y_2)\right)}.
\end{aligned}
$$

Let us note that by construction $J > 0$.
Hence we get, since $(X_1, X_2)^T$ has a bivariate normal distribution $N_2(\mathbf{0}, \Sigma)$, where $\mathbf{0} = (0,0)^T$ that

$$
\begin{aligned}
f_{\mathbf{Y}}(y_1, y_2) &= f_{\mathbf{X}}\left(h_1(y_1,y_2), h_2(y_1,y_2)\right) \mid J \mid \\
&= \frac{1}{2\pi\sqrt{\det\Sigma}}\frac{1}{y_1 y_2\left(1-(y_1+y_2)\right)} e^{-\frac{1}{2}\mathbf{y}^T\Sigma^{-1}\mathbf{y}},
\end{aligned}
$$

where

$$
\mathbf{y}^T = \left(\ln y_1 - \ln\left(1-(y_1+y_2)\right), \ln y_2 - \ln\left(1-(y_1+y_2)\right)\right).
$$

$$
\text{ANSWER}: \underline{f_{\mathbf{Y}}(y_1, y_2) = \tfrac{1}{2\pi\sqrt{\det\Sigma}}\tfrac{1}{y_1 y_2(1-(y_1+y_2))} e^{-\frac{1}{2}\mathbf{y}^T\Sigma^{-1}\mathbf{y}}}, \text{ where}
$$

$$
\underline{\mathbf{y}^T = \left(\ln y_1 - \ln\left(1-(y_1+y_2)\right), \ln y_2 - \ln\left(1-(y_1+y_2)\right)\right).}
$$

Note: The density $f_{\mathbf{Y}}(y_1, y_2)$ is known as (the density of) the *bivariate logistic normal distribution.*

Uppgift 2

(a) The probability generating function of N is by definition

$$g_N(t) = E\left[t^N\right] = \sum_{z=0}^{\infty} t^z p_N(z)$$

$$= p\sum_{z=1}^{\infty} t^z \cdot (1-p)^{z-1} = p\sum_{z=0}^{\infty} t^{z+1} \cdot (1-p)^z$$

$$= pt\sum_{z=0}^{\infty} t^z \cdot (1-p)^z = \frac{pt}{1-t(1-p)}.$$

This requires that $|t(1-p)| < 1$.

ANSWER (a): $\underline{g_N(t) = \frac{pt}{1-t(1-p)}, \quad |t(1-p)| < 1.}$

(b) The characteristic function $\varphi_Y(t)$ of Y is by definition

$$\varphi_Y(t) = E\left[e^{itY}\right] = E\left[e^{itc \cdot S_N}\right] = \varphi_{S_N}(ct),$$

i.e.,

$$\varphi_Y(t) = \varphi_{S_N}(ct), \tag{2}$$

where $\varphi_{S_N}(t)$ is the characteristic function of S_N. We know that

$$\varphi_{S_N}(t) = g_N\left(\varphi_Z(t)\right). \tag{3}$$

where $Z \in L(a)$, $a > 0$ and $\varphi_Z(t)$ is the characteristic function of Z. By the **Formelsamling and Collection of Formulas** we have that

$$\varphi_Z(t) = \frac{1}{1+a^2t^2} \tag{4}$$

When we insert (4) and use the result in part (a) in (3) we get

$$\varphi_{S_N}(t) = g_N\left(\varphi_Z(t)\right) = \frac{p\varphi_Z(t)}{1-\varphi_Z(t)(1-p)}$$

$$= \frac{\frac{p}{1+a^2t^2}}{1-\left(\frac{1-p}{1+a^2t^2}\right)},$$

Thus we get in (2) that

$$\varphi_Y(t) = \varphi_{S_N}(ct) = \frac{\frac{p}{1+a^2(ct)^2}}{1-\left(\frac{1-p}{1+a^2(ct)^2}\right)} = \frac{p}{1+a^2(ct)^2-1+p}$$

and a simple rearrangement yields

$$\varphi_Y(t) = \frac{1}{1+\frac{(ca)^2}{p}t^2}.$$

We see that this suggests that $Y \in L(ac/\sqrt{p})$. It looks like we could take $c = \pm\sqrt{p}$ to get (4). But if it were that $c = -\sqrt{p}$, we would get $Y \in L(-a)$, which is not a permitted Laplace distribution, as $a > 0$. Hence $c = \sqrt{p}$ gives $Y \in L(a)$ by uniqueness of characteristic functions.

This problem describes the Laplace distribution as a *geometrically strictly stable* distribution, as there is a c such that $Y \in L(a)$ if

$$Y \stackrel{d}{=} c \cdot (Z_1 + Z_2 + \ldots + Z_N)$$

where $Z_i \in L(a)$ e.t.c..

ANSWER (b): $\underline{c = \sqrt{p}}$.

Uppgift 3

(a) If $W = \{W(t) \mid 0 \le t < \infty\}$ is a Wiener process, then

$$Y_k = W\left(\frac{k}{n}\right) - W\left(\frac{(k-1)}{n}\right) \in N\left(0, \frac{1}{n}\right)$$

for all k. Thus, as a linear combination of normal random variables,

$$X_n = \sum_{k=1}^{n} e^{-\lambda \frac{k-1}{n}} Y_k$$

is a normal random variable. We need to find its expectation and variance. We have

$$E\left[X_n\right] = \sum_{k=1}^{n} e^{-\lambda \frac{k-1}{n}} E\left[Y_k\right] = 0.$$

and since the increments (i.e., here Y_k) of a Wienerprocess are independent

$$\mathrm{Var}\left(X_n\right) = \sum_{k=1}^{n} e^{-2\lambda \frac{k-1}{n}} \mathrm{Var}\left(Y_k\right) = \frac{1}{n}\sum_{k=1}^{n} e^{-2\lambda \frac{k-1}{n}}.$$

ANSWER (a): $\underline{X_n \in N\left(0, \frac{1}{n}\sum_{k=1}^{n} e^{-2\lambda \frac{k-1}{n}}\right)}$.

(b) By part (a) we have that the characteristic function of X_n is

$$\varphi_{X_n}(t) = e^{-\frac{t^2}{2}\frac{1}{n}\sum_{k=1}^{n} e^{-2\lambda \frac{k-1}{n}}}.$$

We note that the points $\frac{k-1}{n}, k = 1, \ldots n$ form a partition of $[0, 1]$ into n intervals of length $\frac{1}{n}$:

$$0 = \frac{1-1}{n} < \frac{2-1}{n} < \ldots < \frac{k-1}{n} < \frac{k}{n} < \ldots \frac{n-1}{n} < 1.$$

Thus we recognize

$$\frac{1}{n}\sum_{k=1}^{n} e^{-2\lambda\frac{k-1}{n}} = \sum_{k=1}^{n}\frac{1}{n}e^{-2\lambda\frac{k-1}{n}},$$

as a Riemann sum for the Riemann integral $\int_0^1 e^{-2\lambda u}du$. Hence

$$\lim_{n\to\infty}\sum_{k=1}^{n}\frac{1}{n}e^{-2\lambda\frac{k-1}{n}} = \int_0^1 e^{-2\lambda t}dt.$$

Alternatively, i.e., without Riemann sums :

$$\frac{1}{n}\sum_{k=1}^{n} e^{-2\lambda\frac{k-1}{n}} = \frac{1}{n}\sum_{k=0}^{n-1} e^{-2\lambda\frac{k}{n}} = \frac{1}{n}\frac{1-e^{-2\lambda}}{1-e^{-2\lambda\frac{1}{n}}}.$$

We write this as

$$\frac{1}{n}\frac{1-e^{-2\lambda}}{1-e^{-2\lambda\frac{1}{n}}} = \frac{1-e^{-2\lambda}}{\frac{1-e^{-2\lambda\frac{1}{n}}}{\frac{1}{n}}}.$$

Then we set $f(t) = e^{-2\lambda t}$, and recognize the differential ratio

$$\frac{1-e^{-2\lambda\frac{1}{n}}}{\frac{1}{n}} = -\frac{\left(f\left(\frac{1}{n}\right)-f(0)\right)}{\frac{1}{n}} \to -f'(0) = 2\lambda,$$

as $n\to\infty$. Hence

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n} e^{-2\lambda\frac{k-1}{n}} = \frac{1-e^{-2\lambda}}{2\lambda}.$$

We note that $\int_0^1 e^{-2\lambda u}du = \frac{1-e^{-2\lambda}}{2\lambda}$.

Clearly, $\int_0^1 e^{-2\lambda u}du \geq 0$. Hence the preceding entails that

$$\varphi_{X_n}(t) \to \varphi_X(t) = e^{-\frac{t^2}{2}\int_0^1 e^{-2\lambda u}du}, \quad n\to\infty.$$

The limiting function $\varphi_X(t)$ is the characteristic function of $N\left(0, \int_0^1 e^{-2\lambda u}du\right)$. Thus the convergence above implies that

$$X_n \xrightarrow{d} N\left(0, \int_0^1 e^{-2\lambda u}du\right)$$

as $n\to\infty$, as was claimed.

Uppgift 4

Let us note that

$$P\left(X_n \leq n\right) = P\left(X_n - n \leq 0\right) = P\left(\frac{X_n - n}{\sqrt{n}} \leq 0\right).$$

Since $X_n \stackrel{d}{=} U_1 + U_2 + \ldots + U_n$, where U_is are independent Exp(1) distributed random variables and thus $E[U] = 1$, $\mathrm{Var}(U) = 1$. Hence $E[X_n] = n$ and $\mathrm{Var}(X_n) = n$ (see also Appendix 2), and we have by the Central Limit Theorem

$$\frac{X_n - n}{\sqrt{n}} \xrightarrow{d} N(0,1), \quad n\to\infty.$$

Therefore, by the very definition of convergence in distribution,

$$\lim_{n\to\infty} P\left(\frac{X_n - n}{\sqrt{n}} \le 0\right) = \lim_{n\to\infty} F_{\frac{X_n - n}{\sqrt{n}}}(0) = \Phi(0),$$

where $\Phi(x)$ is the distribution function of $N(0,1)$ and as zero is one of its points of continuity (as are all points). But we know that $\Phi(0) = \frac{1}{2}$, and therefore we have established the desired limit as claimed.

Uppgift 5

(a) To demonstrate that

$$\frac{N(n)}{n} \xrightarrow{q} \lambda, \qquad \text{as } n \to \infty$$

we need by definition to show that

$$E\left[\left(\frac{N(n)}{n} - \lambda\right)^2\right], \text{ as } n \to \infty.$$

We expand

$$E\left[\left(\frac{N(n)}{n} - \lambda\right)^2\right] = \frac{E\left[N^2(n)\right]}{n^2} - 2\lambda\frac{E\left[N(n)\right]}{n} + \lambda^2. \tag{5}$$

Since $N = \{N(t) \mid t \ge 0\}$ is a Poisson process with intensity $\lambda > 0$, we have

$$N(n) \in \text{Po}(\lambda n).$$

This yields

$$E\left[N(n)\right] = \lambda n,$$

and

$$E\left[N^2(n)\right] = \text{Var}\left[N(n)\right] + \left(E\left[N(n)\right]\right)^2 = \lambda n + (\lambda n)^2.$$

When we insert these in the right hand side of (5) we obtain

$$E\left[\left(\frac{N(n)}{n} - \lambda\right)^2\right] = \frac{\lambda n + (\lambda n)^2}{n^2} - 2\lambda\frac{\lambda n}{n} + \lambda^2$$

$$= \frac{\lambda}{n} + \lambda^2 - 2\lambda^2 + \lambda^2 = \frac{\lambda}{n},$$

or

$$E\left[\left(\frac{N(n)}{n} - \lambda\right)^2\right] = \frac{\lambda}{n} \to 0 \text{ as } n \to \infty,$$

as was to be proved.

(b) Since $N(n) \in \text{Po}(\lambda n)$, we get by **Collection of Formulas** that the characteristic function of $N(n)$ is

$$\varphi_{N(n)}(t) = e^{n\lambda\left(e^{it}-1\right)}.$$

Thereby the characteristic function of $N(n)/n$ is

$$\varphi_{N(n)/n}(t) = E\left[e^{it\frac{N(n)}{n}}\right] = \varphi_{N(n)}\left(\frac{t}{n}\right) = e^{n\lambda\left(e^{\frac{it}{n}}-1\right)}.$$

Now, as $n \to \infty$,

$$\frac{N(n)}{n} \xrightarrow{q} \lambda \Rightarrow \frac{N(n)}{n} \xrightarrow{P} \lambda \Rightarrow \frac{N(n)}{n} \xrightarrow{d} \lambda$$

But then

$$\frac{N(n)}{n} \xrightarrow{d} \lambda \Rightarrow \varphi_{N(n)/n}(t) \to \varphi_\lambda(t),$$

as $n \to \infty$. Here $\varphi_\lambda(t)$ is a notation for the the characteristic function of λ, which is a discrete stochastic variable with one value $= \lambda$ with probability one. Hence the definition of the characteristic function entails

$$\varphi_\lambda(t) = e^{i\lambda t},$$

and we have shown that the asserted convergence is implied by (1).

(c) We can write

$$N(n) = \sum_{k=1}^{n} Z_k,$$

where

$$Z_k = N(k) - N(k-1) \in \text{Po}(\lambda), Z_1 = N(1) - N(0) = N(1),$$

and Z_ks are independent as the increments of a Poisson process are independent. Also, $E[Z_k] = \lambda$. Thus, we have shown above that

$$\frac{\sum_{k=1}^{n} Z_k}{n} \xrightarrow{P} \lambda,$$

which is an instance of law of large numbers (in the weak sense).

Uppgift 6

We iterate successively and obtain

$$Y_n = Y_{n-1} +_2 X_n = Y_{n-2} +_2 X_{n-1} +_2 X_n = \ldots$$
$$= Y_0 +_2 (X_1 +_2 \ldots +_2 X_n),$$

i.e., .

$$Y_n = Y_0 +_2 (X_1 +_2 \ldots +_2 X_n). \quad (6)$$

We need to compute

$$P(Y_n = 1 \mid Y_0 = 1) = \frac{P(Y_n = 1, Y_0 = 1)}{P(Y_0 = 1)}.$$

But, the rules for addition modulo two entail from (6) that

$$P(Y_n = 1, Y_0 = 1) = P(X_1 +_2 \ldots +_2 X_n = 0, Y_0 = 1)$$

$$= P(X_1 +_2 \ldots +_2 X_n = 0) P(Y_0 = 1),$$

since Y_0 is independent of $X_1, \ldots, X_n$. Therefore

$$P(Y_n = 1 \mid Y_0 = 1) = P(X_1 +_2 \ldots +_2 X_n = 0).$$

The rules for addition modulo two show also that $X_1 +_2 \ldots +_2 X_n = 0$ if only if the sum contains an even number of binary ones. If the number of binary ones is odd, the last (i.e., 'odd') binary one turns the sum $X_1 +_2 \ldots +_2 X_n$ to a binary one. Hence we have

$$P(X_1 +_2 \ldots +_2 X_n = 0) = P\left(\sum_{j=1}^{n} X_j = \text{ even}\right),$$

where $\sum_{j=1}^{n} X_j$ is 'ordinary' addition of zeros and ones. But $W = \sum_{j=1}^{n} X_j$ is a sum of independent Be(p)-random variables, hence $W \in \text{Bin}(n, p)$.